What is it that teachers do that leads to student learning? How can we increase students' achievement and improve their attitudes? In order to improve the quality of education the authors focus on teaching/learning processes taking place in the traditional classroom situations. The study reported is the Dutch contribution to the IEA Classroom Environment Study: Teaching for Learning, whose general objective is to identify alterable teaching behaviors that correlate with desirable cognitive and affective student learning outcomes, and to develop a training program that will recommend ways of organizing teaching/learning processes. Results show that even a short training course can successfully change the teachers' teaching script, and, subsequently their teaching behavior, which in turn influences student achievement.

TS & ZEITLINGER PUBLISHERS

ISBN 90 265 1019-5 NUGI 724

TOMIC & VAN DER SIJDE CHANGING TEACHING FOR BETTER LEARNING

CHANGING TEACHING FOR BETTER LEARNING

W. TOMIC
P.C. VAN DER SIJDE

CHANGING TEACHING FOR BETTER LEARNING

W. TOMIC
P.C. VAN DER SIJDE

SWETS & ZEITLINGER B.V. AMSTERDAM / LISSE PUBLISHERS

Library of Congress Cataloging-in-Publication Data

Tomic, W. (Welko), 1946-
Changing teaching for better learning.

Bibliography: p.
1. Teaching. 2. Teachers--Training of. 3. Learning, Psychology of.
I. Sijde, Peter van der.
II. Title.
LB1025.2.T65 1989 370.1'02 89-11259
ISBN 90-265-1019-5

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, 's GRAVENHAGE

Tomic, W.

Changing teaching for better learning / W. Tomic, P.C. van der Sijde.
Amsterdam [etc.] : Swets & Zeitlinger ; Berwyn : Swets North America.
Met lit. opg.
ISBN 90 265 1019 5
SISO 454 UDC 371.32.02 NUGI 724
Trefw.: onderwijs ; didactiek / lesgeven

Printed in The Netherlands by Offsetdrukkerij Kanters B.V., Alblasserdam
Cover printed in The Netherlands by Casparie, IJsselstein
Cover design by Rob Molthoff

ISBN 90 265 1019 5
NUGI 724

CONTENTS

Chapter 5 DESIGN OF TRAINING PROGRAMS FOR TEACHERS BASED ON THE RESULTS OF RESEARCH *41*

Chapter 6 TEACHER TRAINING AND OBSERVATION INSTRUMENT *75*

Chapter 7 IMPLEMENTATION OF A RESEARCH-BASED TEACHER TRAINING *81*

Chapter 8 THE EFFECT OF A TEACHER TRAINING COURSE ON STUDENT LEARNING OUTCOMES *91*

REFERENCES *105*

PREFACE

Almost everyone nowadays has spent time in classrooms. Even if children leave school at the earliest possible opportunity they will probably have spent over 12,000 hours in various school buildings. This is by far the longest period children will have spent anywhere apart from their own home.
In most countries of the world the predominant form of teaching involves a single teacher engaged in classroom teaching and learning with a group of students varying from about 9 in Denmark and Norway to about 55 in Bangladesh according to the 1978 United Nations Statistical Yearbook.
Naturally, society should be interested in the question what our children learn and how we can increase their achievement and improve their attitudes in particular and contribute to improving the quality of education in general.
The assumption of the study reported in this book is that in order to improve the quality of education we focus on teaching-learning processes taking place in traditional classroom situations.
Both researchers and teachers consider it a great challenge to answer the question what it is that teachers do that leads to student learning. But unfortunately the question is not as simple as it would appear at first glance. To contribute to answering this question, we chose to do research on alterable teaching behaviors in the learning environment. We used the so called "observational-correlational-experimental loop". We have conducted both a correlational and an experimental study in regular classrooms with regular teachers teaching regular subject matter.
The results are neither novel nor earthshaking but are consistent with the outcomes of similar research in this field.
Some parts of the work we are about to report were published in journals. But the present volume is self-contained and includes further research, as well as some rethinking of older formulations and development of new perspectives.
Last but not least, we thank the teachers who voluntarily participated in this study.

Heerlen, Hengelo, The Netherlands — Welko Tomic, Peter Van der Sijde
March, 1989

CHAPTER 1

INTRODUCTION TO THE STUDY

In most countries of the world the predominant form of teaching still involves a single teacher engaged in classroom teaching with a group of about 15 to 50 students.

Although other patterns of organization are possible, it would be difficult to develop any major changes in these well-established institutions and practices. The traditional whole-class, frontal teaching, lecture-recitation-seatwork method predominates because, in our opinion it generally offers the best compromise solution to the dilemmas facing teachers. It enables them to meet more of the needs of more of their students than any feasible alternative method.

Thus more efficient improvements to the education of students in schools across the world must acknowledge the problems associated with major organizational changes and focus on improving the quality of education taking place in traditional classroom situations.

Working on the above assumption, under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) the Classroom Environment Study: Teaching for Learning was conducted that represents a collaborative effort to identify teaching behaviors that affect student learning outcomes and to develop teacher training programs grounded on empirical research. The IEA is a non-government organization of research institutions that co-operate to carry out cross-national empirical educational research.

We would like to outline the aim and the phases of this project briefly in the subsequent paragraphs.

1 THE APPROACH TO EDUCATIONAL CHANGE IN THIS STUDY

Concerning the improvement of education, this study is based on the idea that small changes in a school are easier than large changes. It rests on Popper's notion of "piecemeal engineering" (1972). An important point is that change in teaching behaviors, or aspects of the classroom environment under control of the teacher, can be made at relatively little cost. This kind of change does not ask for reorganization of the school's curriculum, organizational structure, specific buildings or classrooms, nor does it require hiring new staff. The chosen approach in this study is research at the classroom level, the micro-level of education. In our opinion, student learning outcomes will improve substantially.

2 AIMS OF THE STUDY

The Classroom Environment Study in the Netherlands had four aims:
- To identify the teaching behaviors that correlate with desired student learning outcomes in the cognitive and affective domain;
- To examine how far it is possible to design an experimental teacher training program on the basis of the obtained correlational findings that can provide teachers with recommendations about teaching behaviors aimed at improving student learning outcomes;
- To determine the degree to which those teaching behaviors can be fostered through relatively simple teacher training programs;
- To determine the degree to which training and changed behaviors cause improved student achievement and attitudes.

3 PHASES OF THE STUDY

In persuing the aims, the Classroom Environment Study was designed as a three-phase effort. The first phase, the correlational study, attempts to examine the relationships between teaching behaviors and student learning outcomes in order to identify those teaching behaviors correlated with desired student achievement and attitudes. The study observed 50 teachers. Every teacher conducted eight lessons over a period of eight months.

In the second phase, the design of the teacher script, desirable teaching behaviors were formulated using both the obtained results as well as other research results. The assumption was that the results could yield suggestions for desirable teaching behaviors. With regard to a specific teacher training program, a manual was compiled identifying which desired teaching behaviors might affect students' learning outcomes. In other words, the correlational findings were translated into recommendations for possible effective teaching behaviors to be used in designing an experimental teacher training program.

The third phase, the experimental study, first attempts to examine whether a group of teachers will implement the preselected, alterable teaching behaviors in the teacher script. To investigate the influence of the training on teacher competence, a one group pretest-posttest design is used. Second, the effect of the teaching behaviors in the teaching script on student achievement and attitudes has been investigated. Two groups of teachers, namely the experimental group and the control group, participated in a training program. The teachers were placed in the two groups at random.

Ryan and Anderson (1984) published an article on the correlational phase of the Classroom Environment Study. Only in the Netherlands has the complete study involving the three phases been carried out to date.

4 OBSERVATIONAL-CORRELATIONAL-EXPERIMENTAL LOOP

Systematic study of the effects of teaching behavior on student achievement is a relatively new area of research. In recent years, an increasing amount of research has been done on the influences of teaching behavior on student learning. The ultimate objective of such research is to use the teacher to improve the educational process by employing empirical findings regarding the relationship between teaching behavior and student achievements in order to positively influence student learning outcomes.

To reach the above-mentioned four aims, we will make use of the notions of, among others, Rosenshine and Furst (1973) and Gage and Giaconia (1981). According to their view, it is necessary to identify and describe effective teaching behaviors if we plan to develop experimental teacher training courses. The starting-point in identifying effective teaching behaviors are the research results based on the analysis of data gathered in the common practice of teaching. The last description refers to a research program called the observational-correlational-experimental loop, a promising approach in identifying effective teaching behaviors. In short the procedure is as follows. First, one observes and describes the concerned phenomena. One must try to identify the interesting variables in the observed phenomena. Second, correlational studies of those preselected variables must be carried out. Third, one conducts experiments with promising variables from the correlational studies.

This research program is not exclusive to the field of research on teaching, but appears in many other scientific domains. It is used frequently in the medical sciences, for instance in examining the effects of diet on cancer. Recommendations aimed at reducing the incidence of cancers associated with nutrition are based on limited but suggestive evidence from epidemiological studies and experiments with animals (Cohen, 1987). Epidemiologists, like educational re-

searchers, essentially look at the results of experiments in nature and work backward to causes. The most that epidemiologists and educational researchers can do is demonstrate strong and consistent correlations, such as the one between dietary fat and breast cancer, or between positive acknowledgement and desired student achievement. Although statistical associations like correlations can imply causation, they cannot prove it.
On the other hand, researchers conducting experiments work forward from causes to results. On ethical grounds, some laboratory or classroom experiments cannot be conducted upon human beings. Except during the Nazi rule in Germany (Lifton, 1986), medical researchers generally rely on animals. Particular findings are then extrapolated in an effort to assess effects on human beings.
Most research into the influence of teaching behavior on student performance comes from American studies conducted in a nonselective educational system. We agree with Burn and Hurn (1982) that we cannot indiscriminately adopt the results of research focusing on a different educational system. Furthermore, most of the research on this topic has focused on primary education. We cannot assume that results obtained in primary education are applicable to secondary education as well.
The research we have carried out also falls within the "observe- correlate-experiment" pattern.

5 TEACHING BEHAVIORS

The relation between the teaching behaviors observed in this study and the expected student achievements had to be clarified in advance. It was not an easy matter to decide which behaviors we should concentrate upon. We agree with Brophy (1980) and Gage (1979) that behaviors that correlate positively with student outcomes cannot be generalized, meaning that they are not effective for every subject and age level. In selecting the teaching behaviors to be observed, we were guided by the following criteria:
- the teaching behaviors should be manipulable;
- their effects should be observable shortly afterwards;
- they should be well-defined and measurable.

We derived the selected teaching behaviors in the correlational part of the study, first of all, from the variables which Bloom (1976) identified in defining the quality of instruction in his model of school learning. Second, we made use of Levin's literature review (1977), which describes favorable, i.e. effective, teaching behaviors. Third, we based the preselected teaching behaviors on the results of similar studies (Anderson, Evertson & Brophy, 1979; Crawford & Gage, 1977; Fisher, Filby, Marliave, Cahen, Dishaw, Moore & Berliner, 1978; Stallings & Kaskowitz, 1974). All those authors to whom we refer state that the teaching behaviors in their studies are alterable in nature.
By alterable teaching behaviors, we mean those instructional activities related to the presentation or explanation of academic content, such as lecturing/ explaining/ using instructional cues, questioning, allowing students the opportunity to respond, providing feedback, and non-academic activities such as managing students. The term "teaching behaviors" is used interchangeably with the term "interactions" throughout the book.
By student learning outcomes, we mean both student achievement in and attitudes towards mathematics.

6 ORGANIZATION OF THE BOOK

The study reported in this book is the Dutch contribution to the IEA Classroom Environment Study: Teaching for Learning, whose general objective is to identify alterable teaching behaviors that correlate with desirable cognitive and affective student learning outcomes in mathematics. Based on the results of the correlational part of the study, a description of possible effective teaching behaviors will be applied to the development of an experimental training program that will recommend to teachers ways of organizing teaching/learning processes in the classroom to improve students' achievement and attitude scores.
Chapter 2 describes the design of the correlational study and presents the operational variables used.
Chapter 3 reveals of the frequency with which mathematics teachers engage in several behaviors related to classroom interactions and allocation of time during mathematics instruction in eighth grade Dutch classrooms. It deals with an important problem namely, gathering objective evidence on what happens in mathematics classrooms.
Chapter 4 identifies which combination of the different alterable teaching behaviors within the five categories explains most of the variance in student learning outcomes. It discusses suggestive evidence for possible effective teaching behaviors.
Chapter 5 describes the systematic design of an experimental teacher training program. It also discusses a theoretical framework that serves as a reference to the recommendations of the teacher training program.
Chapter 6 provides an outline of the content of the teacher training program followed by a description of the training procedure. In relation to the content of the teacher training program, it also describes the observation instrument.
Chapter 7 describes the effect of changes on the actual teaching behavior as a result of the teacher training program.

Chapter 8 describes the effects of the teacher training program on student learning outcomes.

This study was supported in part by a grant from The Dutch Institute for Educational Research in The Hague and was conducted by the Department of Education, University of Twente in Enschede, and the Department of Social Sciences, Open University in Heerlen.

CHAPTER 2

RESEARCH QUESTIONS AND THE DESIGN OF THE CORRELATIONAL STUDY

We will start this chapter by formulating the research questions of the correlational study. Question 1. How frequently are the preselected alterable teaching behaviors -belonging to the categories teacher lecturing/explaining/providing instructional cues, questioning, giving opportunity to respond, providing feedback, and non-academic interactions- observed during eighth-grade mathematics lessons? The question also examines the context of the observed teaching behaviors and the direction of the interactions: who initiates the interaction and to whom does he or she address him or herself.
Question 2. Which combination of teaching behaviors within the five categories is effective in relation to student learning outcomes in the cognitive and affective domain?

1 RESEARCH DESIGN

To answer the above-mentioned research questions we employed a pre/post-correlational research design. We observed 50 math teachers eight lessons each during an eight-month period.
In order to calculate statistical association measures between teaching behaviors on one side and student learning outcomes on the other, it is in any case necessary to have available data about student learning outcomes at the end of the observation period. However, evidence consistently suggests that cognitive student entry behavior determines to a great extent the results students achieve after having taken a particular course (e.g. Bloom, 1976). Therefore, when investigating influences of teaching behaviors on student learning outcomes, it is necessary to adjust for cognitive entry behavior and affective entry behavior

characteristics of students. To make adjustments regarding influences of students' cognitive entry behavior and affective behavior characteristics in mathematics, pretests are indispensable.

1.1 Population and Sample

Our target population was composed of mathematics teachers working in the eighth grade. All of the seventeen selected schools could be reached by train, and all used one of the following textbooks: "Getal en Ruimte" (Number and Space), "Moderne Wiskunde" (Modern Mathematics), or "Sigma". The schools were drawn at random from schools throughout the Netherlands within the above constraints. The sample involved 50 mathematics teachers.

1.2 Data Collection Regarding Teaching Behaviors

We obtained the data on teaching behaviors or process variables by the direct systematic observation method. We registered the student-teacher interactions using the "Five Minute Interactions Instrument" (FMI). Five well-trained observers registered the applied teaching behaviors in each classroom during five periods of five minutes a lesson totaling eight lessons of approximately 50 minutes. The length of an interaction was five seconds. The FMI instrument was based largely upon instruments developed by the Stanford Research Institute (Stallings, 1977) and is widely used in other classroom studies (Goodlad, 1984). Using the FMI instrument, a survey was obtained of the teaching behaviors during five different five-minute periods of eight lessons. Coding with the FMI instrument was teacher-oriented. Student behavior was coded only when the student interacted directly with the teacher.
Each interaction was specified by means of three codes: the content of the interaction, the context within which the interaction occurred, and who was interacting with whom.

Content
The content area comprised five categories: instruction, teacher questioning, teacher giving opportunity to respond, teacher providing feedback, and non-academic interactions. Instruction included the presentation or explanation of academic content (teacher lecturing/ explaining/ providing instructional cues), with or without material aids. Within the question category a distinction was made between questions requiring short-term memory recall and those requiring more complex cognitive operations, such as the manipulation of information previously acquired. Questions allowing the expression of personal ideas, feelings, or opinions were given their own code. Questions that the teacher redirected to another student following an unsatisfactory response were similarly coded separately.
Student response codes differentiated between responses produced independently and those which involved reading from a text. Any explicit indication that the student did not know the answer was coded as "student does not know".
Within the feedback category, codes specified whether the teacher gave positive or negative acknowledgement to a student response, and noted how frequently the teacher repeated or gave answers. Attempts to seek feedback about teacher effectiveness in explanations or clarifications were also coded.
Absence of teacher/student interactions during academic tasks was coded as "silence".
Non-academic interactions were coded as "discipline", "procedure", or "social".
Whenever background noise prevented the observer from hearing the content of an interaction, the code "cannot hear" was used.

Context
One of six context codes recorded the setting for every interaction. That teachers may interact with varying numbers of students was acknowledged by distinguishing among large group, small group, or private interactions. "Monitoring" was the context code when the teacher supervised students who were working independently. If the teacher was working at his desk and not supervising the students at their seatwork, or if he was interacting with someone outside of the observed class, "non-involved" was indicated. On those occasions when the teacher was observed to be directing the students from one academic activity to another, the "transition" context code was used.

Who to whom
Here the codes identified who initiated the interaction and with whom the interaction occurred. Teacher behaviors were coded as "teacher to group", "teacher to student" in an on-going interaction, or "teacher to others" not within the classroom. Student behaviors were coded as "group to teacher", and "student to teacher".

1.2.1 Training of classroom observers

The five classroom observers were trained for a total of eight days, beginning with six consecutive days before the collection of classroom data, and an additional two days of refresher training three months after the initial training during the period of classroom visits. The training consisted essentially of discussions

familiarizing the observers with the FMI instrument, followed by extensive practice and homework in the use of the instruments. The observers had to learn the definitions of the codes, and become sufficiently proficient so that a minimum of 60 interactions could be coded accurately within a five-minute period. Moreover the training consisted of practical exercises using videotaped and audiotaped classroom situations. All observers visited classrooms for four practice sessions to allow them to practice in a realistic setting.
A more formal check on observer agreement was begun on the third day of training and repeated on subsequent days. Toward the end of each day the trainer compared the observers' frequencies with his own. The research team had set the standard of comparison as a discrepancy no greater than 5 per cent. By the end of the six days, the five observers had achieved an acceptable level of accuracy. Two weeks after the training another check on observer agreement was conducted. These results were also satisfactory.
Although 60 coded interactions per five minutes of classroom instruction initially seems high, this frequency could occur routinely in five-minute periods of classroom time.

1.2.2 Inter-observer agreement and reliability

We examined whether the five observers displayed any differences in coding the teaching behaviors. First, from the data that they collected, we estimated the inter-observer agreement between observer-pairs using the pairwise observation method ("live" observations). Two observers simultaneously performed observations in one classroom. We counted the number of codes in a specific category that each observer used and correlated these counts with one another for each category. We performed this calculation for each of the nine observer-pairs. Second, we estimated the reliability of observation ("audio" observations) by determining the internal consistency coefficient alpha (Cronbach, 1951). To do this, we worked with an additional set of data obtained by having all observers simultaneously code five different five-minute fragments of an audiotaped lesson. A fixed interval of five seconds was set between two codes, which means that the observers were forced to code after each interval of five minutes. Using this interval method, we were able to estimate the coding consistency of all five observers. The context of the teaching behaviors, however, could not be coded from the audiotapes because they lack the essential visual elements.
The results of the reliability checks can be summarized as follows. The mean correlation coefficients calculated on the basis of "live" observations of teaching behaviors vary from -.30 (only one observer-pair) to .97. For context and who-to-whom interactions, the mean correlation coefficients vary from .72 to

.90 and .75 to .85 respectively. The internal consistency coefficients of teaching behaviors calculated on the basis of "audio" observations vary from -.91 (once) to .99. The who-to- whom interactions vary from -1.78 (once) to .99. Note that the coefficients for the context and the who-to-whom interactions are high. To a lesser extent this is also true of teaching behaviors. The above-mentioned results do not deviate from results obtained by similar research (see Sirotnik, 1984).

1.3 Gathering Data on Student Learning Outcomes

By student learning outcomes, we mean cognitive as well as affective learning outcomes. Cognitive outcomes are reflected by the students' scores on an achievement test in mathematics. Affective student outcomes here stand for the students' scores on a scale measuring attitude towards the subject of mathematics.
The pretest was composed of mathematical subject matter for transition year pupils (seventh grade). The pretest consists of test items provided by the Central Institute for Test Development (CITO). For the posttest, the chosen subject matter was treated in the aforementioned eighth-grade textbooks. The posttest consisted of items developed by CITO and ourselves. Data on the variable "students' attitudes towards mathematics" were gathered by means of a Likert-scale.
Regarding cognitive and affective student learning outcomes, we will provide information on the quality of the measurements. We estimated the reliability of the variables in question by testing the internal consistency of the scale (the KR-20 formula) and applying Cronbach's internal consistency coefficient alpha. The KR-20 for the pretest on classroom level is .87. The KR-20 for the posttest on classroom level is .92. The internal coefficient alpha for the pre-attitude and post-attitude scales in mathematics were respectively .60 and .63.

1.4 Operational Definition of Variables

1.4.1 Observed teaching behaviors

A. Teacher lecturing/explaining/providing instructional cues
- Lecture/explain. Presentation or explanation of subject-matter.
- Lecture/explain with materials. Presentation or explanation of subject-matter with material aids.
- Lecture/explain nonverbal. Nonverbal presentation or explanation of subject-matter with material aids, e.g. writing on blackboard or using signs.

- Example. Example of real-life application of mathematical principle.
- Instructional cue. A subject-matter structuring indication, i.e. an indication that structures and facilitates processing of the subject-matter.

B. Teacher questioning
- High-level question. Question requiring student to use complex cognitive operations, such as the manipulation of previously acquired knowledge or information, e.g. reasoning
- Recall or recognition question. Question requiring the student to recall or remember actual information or definitions.
- Directive request action. Question requiring the student to carry out an assignment, e.g. to elaborate a mathematical problem on the blackbord.
- Opinion question. Question allowing the student to express personal ideas, feelings, or opinions on the subject-matter.
- Redirect question. The same question repeated fo another student following an incorrect or unsatisfactory response.
- Probe. Providing the student with additional information following an incorrect or unsatisfactory student response.

C. Opportunity to respond
- Student response. Student response to a question, correct or incorrect.
- Extended student response. Student response to a question, correct or incorrect, longer than ten seconds.
- Student does not know. Negative student answer to question.
- Student statement. Student-initiated task-related statement.

D. Providing feedback
- Acknowledgement positive. Teacher states that the student response is correct.
- Wrong. Teacher states that the student response is wrong.
- Punish. Teacher punishes the student who gives an incorrect response.
- Teacher repeats student answer. Teacher simply repeats student response.
- Teacher answer. Teacher responds following no or incorrect student response.
- Effectiveness of teaching question. Teacher asks student whether he/she understands the subject-matter: feedback for the teacher about his or her explanations or clarifications.

E. Non-academic interactions
- Silence. Absence of interactions during academic tasks.
- Observer can't hear interaction. Background noise preventing the observer from hearing the content of an interaction.
- Discipline. Student misbehavior; teacher as well as student initiated interactions.
- Procedure. Activities reflecting classroom organization e.g. "get out your book", "erase the blackboard", etc.

1.4.2 Composite Teaching Behaviors

A. Lecturing/explaining/instructional cues
C1 All forms of teacher lecture/explain. All forms of teacher explanation, including nonverbal or materially aided, either to the entire class or an individual student.
C2 Lecture/explain to individual student. All forms of teacher explanation, including nonverbal or materially aided, to an individual student.
C3 Instructional cue after lecture/explain. All forms of teacher explanation, to the entire class or an individual student, followed by an instructional cue.

B. Questioning
C4 All forms of questions asked to entire class. All forms of teacher questions to the entire class: high-level and recall or recognition questions, questions related to the effectiveness of explanations or clarifications, redirected questions, probes, and assignments.
C5 All forms of questions asked to individual student. All forms of teacher questions to individual student, identical to C4.
C6 High-level question in large group. Teacher high-level question to group or individual student within a group of more than 8 students.
C7. High-level question in small group. The same as C6, a group of 8 or fewer students.
C8 Recall or recognition question in large group. Teacher recall or recognition question to group or individual student within a group of more than 8 students.
C9 Recall or recognition question in small group. The same as C8, within a group of 8 or fewer students.
C10 Probing group. Teacher probes a question in a large group.
C11 Probing individual student. Teacher probes a question in an individual student.
C12 High-level question followed by probing. Teacher high-level question to group or individual student; if followed by an incorrect student response, teacher subsequently provides feedback and probes the question.
C13 Recall or recognition question followed by probing. Teacher recall or recognition question, etc., the same as C12.

C14 Probing followed by student response. After an incorrect student response to a high-level or recall/ recognition question, the teacher probes the question, after which the same student responds.
C15 Probing followed by correct student response. Teacher probing, followed by student response; the subsequent interaction differs from redirecting, probing, negative acknowledgement, punishment, or teacher response, i.e. the student responds correctly.
C16 Probing followed by incorrect student response. Teacher probing; if followed by an incorrect student response, teacher subsequently provides negative acknowledgement, redirects, or probes.
C17. Teacher waits after asking a question. Teacher asks a question in a large or small group, followed by silence of at least five seconds.
C18 Student cannot answer question. Teacher asks a high-level or recall/recognition question to a group or an individual student, followed by an incorrect student response or silence.
C19 Successful redirecting. Teacher redirects question to another student, whereupon the student responds. The subsequent interaction differs from redirecting, probing, negative acknowledgement, punishment, or teacher response. A correct student response is highly probable.
C20 Redirecting more than once. Teacher redirects question to another student and receives an incorrect student response, the teacher negatively acknowledges this response, and once more redirects the same question.

C. Opportunity to respond
C21 Student gives a very extended response. Student response lasting at least 15 seconds.
C22 Response in chorus. A group of students responds simultaneously.
C23 Student response followed by a new question. Student or a group of students respond, followed by a teacher recall/recognition question or a high-level question.
C24 Student contribution. An individual or class contribution in the form of explanation, reading from a book, or extended response.
C25 Student asks teacher question after lecture/explain. Teacher explains to individual student or class, followed by a student or class question to teacher.
C26 Student asks question. Student high-level or recall/recognition question in a large or small group.
C27 Student asks question or initiates task-related statement. A high-level or recall/recognition question or a task-related statement by an individual student or a group of students.
C28 Teacher response to student question. A high-level or recall/recognition question by an individual student or a group of students, followed by teacher response.

D. Providing feedback
C29 All forms of acknowledgement. The following forms of teacher feedback: positive or negative acknowledgement, punishment, repetition of student answer.
C30 All forms of positive acknowledgement. Student response followed by positive acknowledgement or teacher repetition of student response.
C31 All forms of negative acknowledgement. Student response followed by negative acknowledgement or teacher response.
C32 Teacher criticizes student. Teacher gives negative acknowledgement or punishes student.
C33 No acknowledgement after student response. Student response, possibly incorrect, followed by a non-feedback interaction.

E. Non-academic interactions
C34 Discipline statement to individual student. Teacher discipline statement to individual student.
C35 Discipline statement to entire class. Teacher discipline statement to entire class.

Context of the Teaching behaviors
- Teacher interacting with more than 8 students.
- Teacher interacting with a group of 8 or fewer students.
- Teacher monitoring and helping students during seatwork.
- Transition interactions. Student preparing for a new activity by regrouping, etc.
- Private interactions. Teacher interacting with individual student outside the group.
- Teacher not involved with students. No teacher interaction with students.

Who-to-Whom interactions
- Teacher to group. Teacher addresses him or herself to the entire class or to a group of students.
- Teacher to student. Teacher addresses him or herself to an individual student.
- Student to teacher. Student statement to teacher.
- Group to teacher. Different students address the teacher simultaneously.
- Teacher to others. Teacher interacts with someone other than the student.

CHAPTER 3

TEACHING BEHAVIOR IN MATHEMATICS CLASSROOMS

In this chapter we describe our study of preselected alterable teaching behaviors observed during mathematics lessons. The study was conducted on 50 eighth-grade mathematics teachers at 17 randomly-selected schools. We used the method of systematic observation to record teaching behaviors.
This chapter is organized as follows: we first describe which teaching behaviors we observed during mathematics lessons, i.e., how much time mathematics teachers spend on specific instructional behaviors. Second, we report the context of the observed teaching behaviors. Third, we mention the direction of the interactions, that is, who initiates the interaction and whom he/she addreses. Finally, we discuss whether the results of this study can be used to develop an experimental training course for teachers that might promote the adequate use of allocated time.
To accomplish the above, we observed eight lessons each of 50 math teachers during an eight-month period. Our target population was composed of math teachers working in the eighth grade. All 17 selected schools could be reached by train, and all used one of the following textbooks: "Getal en Ruimte" (Number and Space), "Moderne Wiskunde" (Modern Mathematics), or "Sigma". The schools were drawn at random from schools throughout the Netherlands within the constraints given in the previous sentence.

1 DIRECTLY OBSERVED TEACHING BEHAVIORS

The results regarding teaching behaviors that we gathered using the FMI observation instrument are given below. The results describe the content of the interaction, that is, explanation/instructional cues, questioning, the opportunity for students to respond, feedback, and non-academic teaching behaviors. By interaction, we mean an oral reciprocity between the teacher and students. The mean number of registered interactions per "five minute interaction" is 64.3. After each five seconds the observers coded the interactions. Naturally an interaction could last longer than five seconds, as is the case with the variable "extended student response". When a specific teaching behavior comprises one per cent of the total number of interactions, we observed it approximately 1287 times during the total observation period of 371 lessons. We intended to observe 400 lessons; however, we do not have complete observational data for 29 lessons, in part because of teacher absence, illness, etc.

Table 3.1 gives the results of the observations of teaching behaviors, i.e., the mean frequency of interactions per five minutes, as well as the corresponding standard deviation. The mean frequency of interactions has also been converted to percentages. The total does not amount to 100 per cent because we have not included very small percentages. All categories in table 3.1 are distinct.

Of the registered interactions, 36 per cent pertain to the teacher's lecture/instructional cues (variables 1 through 5). In most cases, the teacher uses the blackboard or the textbook, i.e., material aids, during presentation or explanation of academic content. Only rarely or never does he/she give examples of "real-life application" of mathematical principles. More than 13 per cent of the interactions are instructional cues, remarks that structure the subject matter and enable students to process it better.

Teaching behaviors falling under "questions asked by the teacher" (6 through 11) make up almost 17 per cent of the interactions. Questions that require students to recognize or recall factual information or definitions, i.e., low-level questions, predominate (11 per cent). In comparison, high-level questions or questions requiring more complex cognitive operations, requiring the student to use previously acquired knowledge to find an answer, appear less frequently (3 per cent). The teacher seldom (2 per cent) provides additional information, i.e., probes a question to get the right answer, after a student has given an incomplete or incorrect response. When a question has been answered incorrectly, the teacher very seldom redirects it to another student (0.4 per cent). The percentage of interactions that allow students to express their personal ideas, feelings, or opinions related to mathematical subjects is negligible.

The variables pertaining to the category "give the opportunity to respond" (12 through 15) make up 20 per cent of the total number of interactions. Obviously these observed variables specifically treat student responses. The most frequent response (17 per cent) is an answer to the teacher's question. It is not important, in this connection, to know whether the answer is correct or incorrect.

Table 3.1 Directly observed teaching behaviors

Variable	per 'five minute interaction' mean percentage of interactions	mean frequency of interactions	sd
Teacher lecturing/explaining/ providing instructional cues	36	24	
1 Lecture/explain	6	4.0	2.0
2 Lecture/explain with materials	15	9.6	4.1
3 Lecture/explain nonverbal	2	1.6	1.3
4 Example	-	-	-
5 Instructional cue	13	8.8	2.6
Teacher questioning	16.8	11.4	
6 High-level question	3	2.2	1.7
7 Recall or recognition question	11	7.4	1.9
8 Directive request action	0.4	0.3	0.2
9 Opinion question	-	-	-
10 Redirect question	0.4	0.3	0.3
11 Probe	2	1.2	0.7
Opportunity to respond	20	13	
12 Student response	17	11.1	2.5
13 Extended student response	1	0.5	0.7
14 Student does not know	-	-	-
15 Student statement	2	1.4	0.8
Providing feedback	4.9	3.7	
16 Acknowledgement positive	1	0.8	0.4
17 Wrong	0.5	0.3	0.2
18 Punish	-	-	-
19 Teacher repeats student answer	2	1.6	1.1
20 Teacher answer	0.4	0.3	0.2
21 Effectiveness of teaching question	1	0.7	0.5
Non-academic interactions	20	12.6	
22 Silence	8	5.4	2.1
23 Observer can't hear interaction	2	1.2	1.2
24 Discipline	3	1.6	1.7
25 Procedure	7	4.4	1.6

Of the interactions, 1 per cent are "extended student response", meaning that a student answer lasts longer than ten seconds. We have also included student-initiated, task-related remarks in "give the opportunity to respond"; they constitute 2 per cent of the interactions. Apparently, students hardly ever admit that they don't know the answer to a question. This does not include incorrect

responses that indicate he/she does not know, but is limited to a student's stating that he does not know.
In our study, feedback relates to brief teacher activities, for example a positive or negative acknowledgement of a student response. We did not register feedback immediately followed by a different type of interaction. In other words, we did not include cases where the teacher immediately follows up his/her feedback by probing after a question, repeating a student answer, giving the answer himself, or redirecting the question to another student. Together, the six different forms of feedback (16 through 21) make up 5 per cent of the interactions. According to the results of our study, the feedback concept that the literature (Bloom, 1976 a.o.) emphasizes makes up the category with the smallest percentage of interactions. Within the feedback category, the most-observed variable is "teacher repeats student's answer" (2 per cent). This phenomenon is also called the "pedagogical echo". A teacher indicates that a given answer is correct, i.e., "acknowledgement positive", in no more than 1 per cent of the interactions. He indicates an incorrect answer, i.e., "wrong", to an even lesser extent. Once in a great while the teacher will give the answer to a question himself. The teachers in our sample never punished students for answering incorrectly. We also checked to see whether teachers ask questions related to the effectiveness of instruction, in other words, questions investigating whether the students understand the explanations of the material. Of all interactions, 1 per cent relate to the effectiveness of teaching.
A couple of observed teaching behaviors belong to non-academic interactions (24 and 25). Grouped together they comprise 10 per cent of all interactions. Of these two variables, the most frequent are procedural. Examples of these non-academic interactions are: "get out your book", "erase the blackboard", etc. Relatively few interactions (3 per cent) involve discipline and/or checking students' work.
Finally, two variables constitute the remaining categories: "silence" and "cannot hear" (22 and 23). By "silence" we mean the absence of interactions during the student's academic work. The registered "silence" occurred most frequently while the teacher monitored students as they were engaged in seatwork, and makes up more than 8 per cent of the total number of interactions. Despite the fact that the teachers were armed with a transmitter microphone and the observers with a receiver, in some cases the interactions could not be coded because the observers could not hear the content of the interactions. This was true for 2 per cent of the interactions.

2 COMPOSITE TEACHING BEHAVIORS

In this section we will report the results with regard to the composite teaching behaviors. An important reason for constructing composite variables is to make pronouncements about those teaching behaviors that, while not directly observable, are nonetheless important to the theory of instruction. The values of the composite variables were not calculated directly through observation, the method we used with directly observed variables, but by calculation based on the values of several variables measured through direct observation. Table 3.2 records only the median values, mean values, and standard deviations of the composite variables per five minute interaction. Contrary to the simple counts of directly observed teaching behaviors, the composite teaching behaviors can only be compared in a limited manner with each other. This is particularly true of teaching behaviors that relate to a number of consecutive interactions. On the basis of their mean values, however, we have clearly observed some composite variables more often than others.
Three composite teaching behaviors pertain to "lecture/explain/instructional cues". Not surprisingly, the variable "all forms of teacher lecture/explain" shows a considerably higher mean frequency than "teacher lecture/explain to individual students". This means that most of the lecturing and explaining is directed to the entire group. The least frequent variable in this category is "instructional cue after lecture/explain". The value is small in comparison with the mean value of the directly observed variable "instructional cue" in Table 3.1, a mean frequency of more than 8. Thus, teachers generally do not provide instructional cues immediately after the lecture/explain, but at different moments, for example during task assignments, task performances, and/or seatwork.
Quite a few of the composite teaching behaviors relate to questioning. For ease of survey, we use subcategories within this rather broad category. Thus, we discriminate among questions differing in cognitive level (high-level and recall or recognition questions), questions involving teacher probing, question difficulty, and question redirection. Finally, we have added all forms of questions together in a sum category: "all forms of questions asked by the teacher". This category includes all types of questions posed to individual students and to the entire class. Two composite variables show a quantification of the total number of questions asked in class. The results recorded in Table 3.2 show that teachers ask individual students more questions than they ask the entire class. Comparing this finding with the data of the "lecture/explain" category, we note that "lecture/explain" is usually addressed to the entire group, and that questions are usually addressed to individual students. Composite variables show that recall/recognition questions are more frequent than high-level questions. This

Table 3.2 Composite Teaching Behaviors

Variable	per 'five minute interaction' median	mean frequency	sd
Lecturing/explaining/ instructional cues			
C1 All forms of teacher lecture/ explain	14.0	15.2	4.0
C2 Lecture/explain to individual student	3.0	3.4	1.7
C3 Instructional cue after lecture/ explain	1.8	1.8	0.7
Questioning			
C4 All forms of questions asked to entire class	1.8	2.3	1.4
C5 All forms of questions asked to individual student	6.4	6.9	2.6
C6 High-level question in large group	1.7	2.1	1.5
C7 High-level question in small group	-	0.1	0.3
C8 Recall or recognition question in large group	3.4	3.5	1.5
C9 Recall or recognition question in small group	0.8	0.8	0.6
C10 Probing group	-	0.1	0.1
C11 Probing individual student	1.0	1.2	0.6
C12 High-level question followed by probing	0.2	0.3	0.3
C13 Recall or recognition question followed by probing	0.3	0.4	0.2
C14 Probing followed by student response	0.1	0.1	0.1
C15 Probing followed by correct student response	0.6	0.6	0.3
C16 Probing followed by incorrect student response	0.3	0.4	0.2
C17 Teacher waits after asking a question	0.9	1.0	0.5
C18 Student cannot answer question	0.6	0.6	0.3
C19 Succesful redirecting	0.2	0.2	0.1
C20 Redirecting more than once	-	-	-
Opportunity to respond			
C21 Student gives a very extended response	0.9	0.3	0.3
C22 Response in chorus	1.1	1.4	1.0
C23 Student response followed by a new question	1.3	1.6	0.9
C24 Student contribution	1.4	1.7	1.0

Table 3.2 continued

Variable	median	mean frequency	sd
C25 Student asks question to teacher after lecture/explain	0.5	0.5	0.3
C26 Student asks question	1.5	1.6	0.7
C27 Student asks question or initiates task-related statement	4.2	4.4	2.2
C28 Teacher response to student question	1.7	1.8	1.1
Providing feedback			
C29 All forms of acknowledgement	3.8	4.0	1.8
C30 All forms of positive acknowledgement	1.4	1.6	0.8
C31 All forms of negative acknowledgement	0.3	0.3	0.2
C32 Teacher criticizes student	0.2	0.3	0.2
C33 No acknowledgement after student response	3.9	4.2	1.4
Non-academic interactions			
C34 Discipline statement to individual student	0.6	1.1	1.2
C35 Discipline statement to entire class	0.2	0.4	0.4

also applied to directly observed teaching behaviors. The subcategory "probing" includes seven composite variables. Examining these consecutive interaction combinations, we see that a specific sequence is more frequent with regard to individual students than to the large group. The difference between the mean values of the variables "recall or recognition question followed by probing" and "high-level question followed by probing" is small, considering that recall/recognition questions are more frequent. When a teacher probes after a question, the student will more often give a correct response than an incorrect one. We have rarely observed the remaining composite teacher behaviors in this subcategory. The table also records the results of the composite teaching behaviors that indicate the difficulty level of the questions the teacher intends to ask. The most frequent is the combination of interactions in which the teacher waits at least five seconds after asking a question before calling on a student. Apparently, students do not respond correctly to a recall/recognition question or high-level question in every case. Of the total number of questions asked, adding C4 and C5, more than 6 per cent of the questions appear to be too difficult for students. Table 3.2 shows that teachers redirect a question to another student who is likely to give the right answer, though relatively infrequently. Comparing this finding with the directly observed variable "redirecting", we see that a redirected question leads to a correct student response in two thirds of all cases. A teacher rarely redirects a question more than once.

Naturally, in the category "give students the opportunity to respond/make contributions", we primarily recorded student behaviors. The mean value of 1.3 belongs to the composite variable "student response followed by a new question". This means that the teacher asks another recall/recognition or high-level question immediately after the students have responded. Furthermore, our results show that students do respond in chorus to a teacher question. Students rarely give very extended responses, i.e., answers lasting longer than 15 seconds. The results also tell us that students will take the initiative in asking questions or making task-related statements during mathematics lessons. To a lesser extent, it appears from the results that teachers answer questions asked by the students. If we consider the low mean value of the corresponding composite variable, students apparently do not often ask a question after the teacher's lecture/explain. In any case, the data show that students are allowed to demonstrate initiative and to contribute during math lessons.

In the category "feedback", note that the calculated mean value is highest for the composite teaching behaviors "no acknowledgement after student response", even when this response is incorrect. By a small margin, "no acknowledgement after student response" is even more frequent than "all forms of acknowledgement provided by the teacher". We can also see that "all forms of positive acknowledgement" are more frequent than "all forms of negative acknowledgement".

Finally, two composite teaching behaviors in table 3.2 relate to disciplinary statements. We call these non-academic interactions. Generally, interactions that qualify as discipline statements rarely take place during mathematics lessons. Teachers more frequently direct discipline statements towards individual students than to the entire class.

3 CONTEXT OF THE TEACHING BEHAVIORS

The results regarding the context of teaching behaviors registered in Table 3.3 show that most of the interactions take place between the teacher and a large group of students (77 per cent). More than 20 per cent of all interactions occur while the teacher moves about the classroom, checking the work of individual students and helping students working independently or in subgroups. In more than 1 per cent of all interactions, the teacher is at work at his desk and consequently not involved with the students. Other context variables, such as transition and private interactions with a student, are rare.

Table 3.3 Context of the Teaching Behaviors

Variable	per 'five minute interaction' mean percentage	median	mean frequency	sd
Teacher interacting with more than 8 students	77	50.9	50.4	8.8
Teacher interacting with a group of 8 or fewer students	1	-	0.8	2.5
Teacher monitoring	20	11.4	13.3	7.8
Transition interactions	0.3	-	0.2	0.4
Private interactions	0.5	-	0.4	0.7
Teacher not involved with students	1.3	0.7	0.8	0.9

4 WHO-TO-WHOM-INTERACTIONS

With this part of the FMI observation instrument, we have registered who initiated the interaction and with whom the interaction occurred. Table 3.4 differentiates five different situations.

Table 3.4 Who-to-Whom Interactions

Variable	per 'five minute interaction' mean percentage	median	mean frequency	sd
Teacher to group	43	27.4	28.2	5.2
Teacher to student	31	19.7	20.1	3.9
Student to teacher	21	13.8	13.8	2.7
Group to teacher	4	2.1	2.4	1.3
Teacher to others	1	0.8	0.9	1.0

The results recorded in the table above show that the teacher initiates 74 per cent of the interactions. In 43 per cent of the interactions, the teacher addresses the entire class or a large group of students. In a considerable number of interactions, almost 31 per cent, the teacher addresses a single student. This personal approach to the student differs from what educators call individualization. Although the teacher primarily initiates interactions, students may also do so. The student addresses the teacher in 21 per cent of all interactions, and in some cases the entire group of students addresses the teacher.

5 DISCUSSION

This chapter described the instructional behaviors of mathematics teachers. We investigated, in other words, how teachers used the allocated time for mathematics instruction. This is of great importance; recent years have produced strong indications that efficient use of allocated time correlates positively with student learning outcomes (Bloom, 1977, a.o.). Because instructional behaviors are thought to be more or less general (Brophy, 1979), our summary and conclusion will, as far as possible, compare our results with the findings of a study conducted by Goodlad (1984) in the USA. The observation system he used was almost identical to the one discussed in this chapter. Next, we will provide potential explanations for a number of results. Finally, we will see to what extent we can use the results derived from this study to develop an experimental teacher training course.

Almost all interactions (97 per cent) took place either during frontal instruction, while the teacher was checking the work of individual students, or while helping them as they worked independently. We found that in many cases, teachers addressed the entire class. We also found that a considerable percentage of the interactions (31 per cent) took place between the teacher and individual students. The teacher initiated almost three-fourths of all interactions. Since 25 per cent of the interactions were initiated by students, we can conclude that they were given the opportunity to initiate interactions themselves. Apparently, it is not unusual in mathematics instruction for students to make a contribution or show initiative during a lesson. The predominant organization within the class was the large workgroup. The teacher usually communicated with the entire class. Teachers rarely instructed small groups. Inevitably, we must conclude that frontal instruction predominates. Goodlad (1984) reached the same conclusion on the basis of his findings. The dominant role of the teacher regarding the course of instruction is also evident. He is the central figure in the classroom. This finding also corresponds to that of Goodlad (1984).

Concerning the content of the interactions, i.e., teaching behaviors related to instruction, interactions pertaining to the presentation of subject matter were not the most frequent (36 per cent). Two-thirds of these interactions were part of the "lecture/ explain" category. The only material aids during "lecture/ explain" are the textbook and the blackboard. Our results show that, during lecture, teachers rarely gave examples of the daily use of mathematical principles. Apparently, math teachers do not consider the application of mathematical principles a priority in the second year of secondary education. This corresponds with Krammer's findings (1984) that teachers rarely chose practical problems from the mathematics course they used.

Almost 17 per cent of the interactions were related to teaching behaviors involving "asking questions". Over 3 per cent of all interactions were questions requiring more complex thought processes. Teachers mainly asked low-level questions requiring the recall or recognition of information. They more frequently addressed questions to individual students than to the entire class. This is also true in the few cases when teachers probed after questions. When students gave an incorrect or incomplete answer to a question, the teacher tended to probe the same student - enabling him/her to improve or complete his or her answer - rather than ask another student the same question.

Almost 20 per cent of the interactions were part of the category "give the opportunity to respond". Within this category, the most frequent variable was the response the student gave to a teacher's question.

Over 5 per cent of the teaching behaviors can be considered "feedback". Teachers more often provide positive than negative feedback. They do not punish students for giving an incorrect answer.

More than 10 per cent of the interactions were non-academic behavior. Within this category, procedures predominate. Relatively few of the observed teaching behaviors were of a disciplinary nature.

To end this summary, we note that from our collected data we can derive patterns of instructional activities that fairly accurately describe the procedures taking place during mathematics lessons.

An important point related to questions and to a lesser extent to other teaching behaviors, is the quality issue. Besides knowing how often questions of a certain cognitive level appear in a lesson, it is also important to know how good the questions are and whether they are asked at appropriate times. However, this quality issue is beyond the scope of this study.

Assuming that we have accurately described the 50 classes examined in this section, in what way can we explain the teaching behaviors observed? The results have a number of potential explanations. To explain the dominance of frontal instruction, the dominant role of the teacher, and his demonstrated initiative, we first place great value on Goodlad's hypothesis (1983): teachers teach in the same way they themselves were taught. They use the same behaviors and techniques they were confronted with in their long careers as students (about 18 years for the highest qualified teacher in the Netherlands). Apparently, the lengthy modeling process was a greater influence than the short period of vocational training. Second, the traditional whole class, frontal teaching predominates, probably because it still is generally the best compromise solution to the dilemmas facing teachers. The popularity of this form of instruction lies in part in the lack of convincing evidence that other forms would be more efficient, i.e., would enable teachers to meet more of the needs of more of their students. According to Jamison, Suppes and Wells (1974), attempts to replace frontal teaching, e.g., by individual instruction with the assistance of media such as programmed instruction and computer-assisted instruction, did not demonstrate results justifying the "abolition" of frontal teaching.

We have seen that teachers rarely gave examples of the use of mathematical principles in daily life. We may be able to explain this by noting that few problems on math examinations show a relation to applied, every-day mathematics.

Although teachers asked recall/recognition questions more frequently than high-level questions, we have seen that they probed after the latter more often. If we assume that teachers only probe after questions when a student has given an incorrect answer or has not understood the question, then probing wil most likely occur after the teacher has posed a question requiring a high cognitive-level answer.

Concerning the difficulty of the questions, we found that teachers waited at least five seconds between asking a question and calling on a student. Possibly teachers considered such questions difficult and therefore gave their students the opportunity to think before responding.

Our results show that students sometimes respond to a question in chorus. We do not know whether the teacher has asked them to do so, since we have no information on this specific variable, but perhaps this indicates the degree of the students' enthusiasm.

We have also seen that students hardly ever asked questions after lecture. We might consider this a positive qualification of the lecture. Perhaps Good, Slavings, Harel, and Emerson (1987) give a more plausible explanation for this student-initiated behavior when they state that students learn in subtle ways to avoid asking questions rather than risk indicating that they do not understand the teacher's lecture.

As far as teacher feedback is concerned, this category included the smallest number of interactions. We can explain this small percentage, surely underestimated by the way we decided to register this teaching behavior. In the first place, our study related feedback to brief teacher activities. In the second place, we did not register feedback immediately followed by an interaction of another category. Naturally the 50 classrooms experienced teacher feedback more often than our results suggest.

Our results also show that teachers seldom indicate when a student's response is incorrect. We may explain this from the teacher's attitude; he is more likely to reward than to disapprove. Also, since all forms of positive acknowledgement are more frequent than all forms of negative acknowledgement, in our opinion teachers are more likely to reward correct answers than to negatively acknowledge incorrect answers. Students' responses are more frequently correct than incorrect, partly because recall or recognition questions are more frequent than high-level questions. Though the cognitive level of a question is conceptually separate from its difficulty level (Brophy & Good, 1986), we think that recall/recognition questions in the context of eighth grade mathematics are less difficult to answer than high-level questions.

We did not observe teachers punishing students for giving an incorrect answer; this, and the above-mentioned remark regarding the teacher's attitude, may first be ascribed to the so-called "observation effect"; the teacher probably behaves more prudently and checks him/herself in the presence of another adult observing his/her work in the classroom. Second, teachers have little reason to punish students who give incorrect answers, beyond merely providing negative feedback, e.g., in the form of personal criticism.

The results described in this chapter concern the instructional behaviors of teachers related to the use of allocated time. The question is whether these results can serve as the starting point in formulating and developing an experimental teacher training course that will promote the appropriate use of allocated time. It goes without saying that this question is difficult to answer. Even if the observed behaviors correspond with desired student performances -and evidence indicates this (Levin, 1977)- we still could not conclude that teachers should frequently apply these teaching behaviors. Perhaps these teaching behaviors have to be applied in a specific way and at specific moments in the instructional process. As a matter of course it is important to know how many times a teacher engages in a certain type of instructional action and how many times certain behaviors occur, but in our view, the instructional message that the action carries is important as well. An important question is whether the teacher conciously chooses specific instructional behaviors over others, and why. However, we do have one point of departure for developing an experimental training course for teachers. The results show that 10 per cent of the allocated time was spent on activities that had nothing whatsoever to do with mathematics, i.e., non-academic activities. These behaviors mainly involved procedures. In Tomic (1983), we reported that 68 per cent of the classes participating in the sample received three hours of mathematics instruction per week. Assuming that a year consists of 35 schoolweeks, an average class has 105 lessons in mathematics. Averaged annually, the above-mentioned 10 per cent of the allocated time yields almost 11 math lessons spent on non- academic activities. Although some time spent on procedures is inevitable in any classroom, an experimental training course that trains teachers to focus on student learning and student management might help reduce the time used for non-academic activities. If such training included instruction in effective teaching behaviors on the basis of research results, student learning outcomes might possibly improve (Van der Sijde & Tomic, 1987).

CHAPTER 4

TEACHING BEHAVIOR AND STUDENT LEARNING OUTCOMES IN MATHEMATICS CLASSROOMS

In all educational activities, teachers try to realize specific learning objectives. They do this by applying more or less specific teaching behaviors.
In this chapter we will describe our study identifying those teaching behaviors that have cognitive and affective learning effects on students taking a course in mathematics.
We will examine which combination of teaching behaviors within the five different categories best predicts student learning outcomes, as measured by the average class posttest scores in the sample of 50 teachers. We determined the multiple regressions of mathematics achievements and the attitude of students towards mathematics for each category of teaching behaviors. In this way, we traced which linear combination of teaching behaviors explains most of the variance in student learning outcomes. We distinguish between teaching behaviors affecting student learning outcomes both in the cognitive and affective domains.
Once we have successfully described effective teaching behaviors, we expect to design an experimental teacher training program on the basis of this material.

1 METHOD OF ANALYSIS, MULTICOLLINEARITY AND SELECTION PROCEDURES

To address the problem of multicollinearity, we followed the suggestions made by Neter and Wasserman (1974). On theoretical grounds, we first reduced the number of teaching behaviors. Second, we inspected the intercorrelation matrix of the teaching behaviors in a particular model. Third, we removed the inde-

pendent variables that produced multicollinearity in the model.
To avoid multicollinearity, several independent variables were not submitted to the analysis. We selected the teaching behaviors to be analysed on conceptual grounds and on their suitability for an experimental teacher training program. As a result of the two selection criteria, almost all the reported teaching behaviors are composite variables.
To reduce the number of independent variables, we applied the backward procedure, entering the class mean scores on the cognitive, respectively affective pretest as covariables for all cases. We know that entering behavior explains a great part of the variance in learning outcomes (Bloom, 1976, a.o.). Naturally, we must adjust for this confounding variable. In our study, the cognitive pretest scores explain 43 per cent of the variance in mathematics achievement and the affective pretest scores explain 13 per cent of the variance in the affective posttest scores.
The results of the multiple regressions of the criterion variables on the independent variables are found in table 4.1 and 4.2. We exclude those cases where the teaching behaviors, in addition to scores on the cognitive pretest, explain less than 2 per cent of the variance in mathematics achievement, respectively students' attitudes towards mathematics. The tables report the following information: the regression coefficient (B), standard error B, the t-ratio, the significance level, the standardized regression coefficient (Beta), the multiple correlation coefficient R, and the determination coefficient R^2.
In the following, we present the results of the multiple regressions of mathematics achievement, respectively students' attitudes towards mathematics, for different teaching behaviors within the distinct categories.

2 RESULTS IN THE COGNITIVE DOMAIN

After the multicollinearity check three teaching behaviors remained that explain less than 2 per cent of the variance in mathematics achievement in addition to cognitive entering behavior: "all forms of teacher lecture/explain", "lecture/ explain to individual student", and "instructional cue after lecture/explain". Table 4.1 shows no results for behaviors that fall within the category "opportunity to respond", since they explain less than 1 per cent of the variance in mathematics achievement.

Cognitive level of questions Originally, the linear regression model consisted of the following variables: "student cognitive entering behavior", "recall/recognition question", "high-level question in large group" and "recall/recognition question in large group". Together, the four predictors explained 46 per cent of

the variance in mathematics achievement. After using the backward procedure, in addition to "student cognitive entering behavior", only the teaching behavior "high-level question in large group" remained in the model. Combined with "student cognitive entering behavior", this teaching behavior explains 45 per cent of the variance in mathematics achievement. This means that high-level questions in large groups describe 2 per cent of the variance in mathematics achievement.
Probing The model included seven predictor variables that, taken together, explained 46 per cent of the variance in mathematics achievement. In addition to "student cognitive entry behavior", the model treated "probing to group", "high-level question followed by probing", "recall/recognition question followed by probing", "probing followed by student response", "probing followed by correct student response" and "probing followed by incorrect student response". After analysis, two teaching behaviors remained, namely "probing followed by student response", and "probing followed by correct student response".
Difficulty level of questions The analysis of this model involved two predictor variables, namely "student cognitive entering behavior" and "teacher waits after asking a question", indicating the difficulty level of the teacher's question. In this model, this teaching behavior apparently explains 2 per cent of the variance in mathematics achievement.
Redirecting In addition to "student cognitive entering behavior", this model included the predictors "redirect question", and "successful redirecting". After we applied the backward procedure, these two teaching behaviors remained in the model and explain 4 per cent of the variance in mathematics achievement.
Student contribution Originally the model was composed of the predictor variables "student contribution", "teacher response to student question", and "student asks question", in addition to "student cognitive entry behavior". By applying the backward procedure we eliminated two teaching behaviors and the variable "student asks a question" remained in the model. The standardized regression coefficient of this predictor variable is negative.
Providing feedback Teaching behaviors in the model belonging to the category "providing feedback" are: "teacher answer", "all forms of positive acknowledgement", "all forms of negative acknowledgement", and "no acknowledgement after student response". After analysis, the behaviors "all forms of positive acknowledgement" and "all forms of negative acknowledgement" remained in the model and explain 4 per cent of the variance in mathematics achievement.
Non-academic interactions In addition to "student cognitive entry behavior", three teaching behaviors entered the analysis, namely "procedural interactions", "discipline statement to individual student", and "discipline statement

to entire class". After analysis, only the teaching behavior "discipline statement to individual student" remains in the model and explains 4 per cent of the variance in mathematics achievement.

Table 4.1 Results of Backward Multiple Regression Analysis of Mathematics Achievement on Cognitive Entry Behavior and Teaching Behaviors

Variables	B	Standard error B	t	sig	Beta	Mult R	R^2	F
model 1								
C6 High-level question in large group	.25	.18	1.37	.17	.15	.67	.45	19.57*
model 2								
C14 Probing followed by student response	-4.86	4.23	-1.15	.25	-.1	.67	.45	12.82*
C15 Probing followed by correct student response	1.30	1.01	1.29	.20	.17			
model 3								
C17 Teacher waits after asking a question	.61	.51	1.19	.23	.12	.67	.45	19.16*
model 4								
10 Redirect question	-.98	1.06	-.92	.35	-.10	.69	.47	13.67*
C19 Successful redirecting	3.09	1.94	1.59	.11	.17			
model 5								
C26 Student asks question	-.66	.38	-1.75	.08	-.18	.68	.47	20.61*
model 6								
C30 All forms of positive acknowledgement	.74	.39	1.86	.06	.25	.69	.47	13.73*
C31 All forms of negative acknowledgement	-1.94	1.59	-1.22	.22	-.16			
model 7								
C34 Discipline statement to individual student	-.41	.22	-1.84	.07	-.19	.69	.47	20.91*

*p < .001

3 RESULTS IN THE AFFECTIVE DOMAIN

Table 4.2 does not report results on teaching behaviors belonging to three categories, namely the cognitive level of questions, the difficulty of questions and, student contribution; in addition to affective entry characteristics, the explained variance in the affective posttest scores for these categories is less than 2 per cent.

Lecturing/explaining/providing instructional cues In the model, the teaching behaviors "all forms of teacher lecture/ explain", "lecture/explain to individual student", and "instructional cue after lecture/explain" entered the analysis. In combination with "student affective entry characteristics", these predictor variables explained 16 per cent of the variance in the affective posttest scores. After we applied the backward procedure, one teaching behavior, "all forms of teacher lecture/explain", remained in the model and explains 2 per cent of the variance in the criterion variable.

Table 4.2 Results of Backward Multiple Regression Analysis of Mathematics Attitude on Affective Entry Characteristics and Teaching Behaviors

Variables	B	Standard error B	t	sig	Beta	Mult R	R^2	F
model 1								
C1 All forms of teacher lecture/explain	.03	.03	.96	.33	.13			
model 2								
C16 Probing followed by incorrect student response	-1.17	1.05	-1.11	.26	-.24	.48	.23	2.71*
C10 Probing to group	-2.99	1.79	-1.67	.10	-.25			
C12 High-level question followed by probing	-.83	.95	-.87	.38	-.18			
C15 Probing followed by correct student response	.43	.93	1.52	.13	.39			
model 3								
C30 All forms of positive acknowledgement	.27	.17	1.50	.13	.19	.49	.24	4.85*
20 Teacher answer	-1.37	.66	-2.06	.04	-.27			
model 4								
C34 Discipline statement to individual student	-.13	.13	-1.00	.32	-.13	.39	.15	4.25*

*p < 0.5

Probing The model was composed of seven predictor variables. The six teaching behaviors are the same as those in the probing model in the cognitive domain. Four teaching behaviors remained in the model. "Probing followed by

an incorrect student response", "probing to group", and "high-level question followed by probing" have a negative relationship with students' attitudes towards mathematics. "Probing followed by correct student response" is positively related to attitudes towards mathematics in the model.
Providing feedback Five teaching behaviors related to feedback were included in the model, namely "acknowledgement positive", "wrong", "teacher repeats student answer", "teacher answer", and "effectiveness of teaching question". After analysis, two teaching behaviors explaining 11 per cent of the variance in students' attitudes towards mathematics remained in the model. Those teaching behaviors are "all forms of positive acknowledgement" and "teacher answer". The regression coefficient of the latter variable is negative.
Non-academic interactions This model included the same variables as the model in the cognitive domain. After analysis, only one teaching behavior, "discipline statement to individual student", remained in the model and has a negative effect on students' attitudes towards mathematics.

4 DISCUSSION

In this chapter we investigated which combination of the different teaching behaviors was most effective in predicting student learning outcomes in the cognitive and affective domain. We determined the multiple regressions of mathematics achievements and the attitude of students towards mathematics for each category of teaching behaviors. In this way, we traced which linear combination of teaching behaviors explains most of the variance in student learning outcomes. Statements about relationships between teaching behaviors on the one side and student learning outcomes on the other indicate to what extent teaching behaviors can predict learning outcomes. Because student entering behavior is the most important predictor of student learning outcomes, we entered "student cognitive entering behavior", respectively "student affective entry characteristics" as the first variable in all cases in the regression model. In our study, the scores on the cognitive pretest explain 43 per cent of the variance in mathematics achievement, and scores on the affective pretest explain 13 per cent of the variance in scores on the affective posttest.
We will summarize the results concerning effective teaching behaviors, comparing our outcomes, as far as possible, with results and/or theoretical points of view of other researchers. Since most researchers have reported about learning outcomes in the cognitive domain, we will restrict this comparison mainly to that domain. After this, we will mention some implications for the design of an experimental teacher training program.
As mentioned before in this chapter, for the most part, we report only composite teaching behaviors. In our opinion, patterns of teacher-student interactions are more interesting and significant for designing teacher training programs than directly observed, discrete teaching behaviors.

In the category of teaching behaviors related to lecture/explain/instructional cues, we observed no effect on mathematics achievement. Other researchers report positive effects on achievement (Anderson et al. 1979; Bloom, 1976; Clark et al. 1979; Fisher et al. 1978; Levin, 1979; Lysakowski & Walberg, 1982). We did find a positive effect in the affective domain. This means that the more time teachers spend on lecture/explain, the more positive the students' attitude towards mathematics will be. Taking into account that teaching behaviors related to lecture/explain make up 36 per cent of classtime, the explained variance is small. Maybe students find this rather passive activity of listening pleasant. Another explanation is that the variable "enthusiasm", a predictor of learning outcomes (Rosenshine & Furst, 1971), leads to frequent lecturing and explaining.
Within the category "questioning", high-level questions had a positive effect on mathematics achievement. The same effect is mentioned by Levin (1977) and Redfield and Rousseau (1980) among others. In our study, we observe that questions put to a large group of students are more effective than questions put to a small group of students. Concerning the difficulty of questions, there appears to be a positive effect on mathematics achievement when a teacher waits after asking a question. Obviously, inserting a pause after asking a question contributes to learning outcomes. Maybe a correct response is more likely when students get to consider their answers carefully. In general, the frequency of teaching behaviors related to probing and redirecting was rather low. Nevertheless, we found redirecting a question had a positive effect on mathematics achievement.
No positive effect on student learning outcomes was found for behaviors within the category "opportunity to respond". In our study, student-initiated, task-related statements have a negative effect on learning outcomes. The same holds for the behavior "student asks a question" on mathematics achievement. This contradicts the results reported by Brophy and Evertson (1974) and Good and Grouws (1975).
In view of the strong theory and the amount of research that has been carried out to date (Bloom, 1976; Fisher et al., 1978; Lysakowski & Walberg, 1981, 1982), we were not surprized to find that "all forms of positive acknowledgement" have a positive effect both on mathematics achievement and attitude towards mathematics. Naturally, we observed a negative effect for "all forms of negative acknowledgement". Again, positive acknowledgement is apparently an effective teaching behavior, while negative acknowledgement does not contribute to learning outcomes.

On the subject of non-academic interactions, we found that the more time spent on monitoring and management of learners, the worse the learning outcomes. One obvious explanation is that teachers manifesting inadequate management behavior spend a lot of time making discipline statements; consequently, students learn very little.

Regarding the influence on the future entry behavior of students, we think, on the basis of our results concerning the explained variance, that in principle teachers can change more in students' attitude towards mathematics than they can contribute to students' knowledge of mathematics. This conclusion seems important for teaching the subject, for one assumes that the more positive the attitude of students towards mathematics, the higher they score on cognitive tests.

We agree with Evertson, Emmer and Brophy (1980) that the large number of variables related to effective teaching behavior show its multiple facets. It cannot be reduced to a single behavior. Since we analysed teaching behaviors within behavior categories, we cannot say which of all observed behaviors is most important in producing achievement or positive student attitudes. The applied analysis method was therefore inappropriate.

Although these results have been observed in other process-product research (Rosenshine and Furst, 1973; Good and Grouws, 1977; Evertson, Emmer and Brophy, 1980) our study found strikingly few effective behaviors. We ought to consider that, first of all, the student population had above-average achievement levels compared with other twelve-to-thirteen-year-olds in the Netherlands, and second, that our study was conducted in a different educational system, namely in highly selective schools.

We have seen that more teaching behaviors have a positive effect on mathematical knowledge than have a positive effect on students' attitude towards mathematics. To appreciate the different teaching behaviors in the cognitive and affective domain, we should recognize their similarities.

Perhaps a training course focusing on learning- and student- management, will help to reduce the percentage of inactive students, as well as the time spent on non-academic interactions. However, we must not interpret these results causally. We cannot identify which teacher behaviors are effective without conducting further research. Fenstermacher (1976) also pointed this out. We cannot confine ourselves to the results of studies linking teaching behavior and student outcomes, and we agree with Gilbert and Mosteller (1972) that strictly controlled, ecologically valid field experiments are necessary to determine effectiveness. We must discover whether there is a causal connection between some teaching behaviors and a student's success at a given task. Only if the answer is positive, can we recommend teaching behaviors that positively affect learning outcomes,

and with important implications for pre- and in-service teacher training.

CHAPTER 5

DESIGN OF TRAINING PROGRAMS FOR TEACHERS BASED ON THE RESULTS OF RESEARCH

In this chapter we subscribe to the view that a contribution to the professionalization of teacher trainings can be made by reserving a part of the training program for teaching behaviors that promote desireable student learning outcomes. Studies in which research is undertaken into the effect of teaching behaviors on student learning outcomes are carried out according to the so-called "observational-correlational-experimental loop". As we wish to emphasize teacher training programs, we will describe four model studies in which such programs occupy a central position. The results of these studies show that when teachers are trained in certain behaviors, their students score higher on a cognitive achievement test. As we make a plea for the systematic design of teacher training programs, we describe a model for such a design. We use that model to analyse the design process of the training programs in the four model studies. The analysis shows that information about the design and development process is generally limited. With regard to two phases of the design process, that is the design of the training program and the evaluation in the sense of "teacher competence" almost all information is lacking.

The most common form of instruction, a teacher teaching a group of about 25 students, appears unlikely to change in the near future. This form is popular in part because there is little convincing evidence that other forms are more efficient. This has already been established by Jamison, Suppes, and Wells (1974). Attempts to replace this form of classroom teaching by individual instruction with the assistance of such media as programmed and computer-assisted instruction did not appear to show results that justify the "abolition" of classroom teaching. Ten years later, Goodlad (1984), in his comprehensive study of more than 1000 schools concluded that classroom teaching is the most common form of instruction. Cuban (1984), in his study on the history of teaching in America, concurs. Given these points, it is reasonable in our view to take classroom teaching as a starting point for the study of teacher effectiveness.

We think it is a good premise in connection with classroom teaching to regard the teacher as an intermediary who teaches the subject matter to a group of students with the aim of improving student outcomes.

There are two research approaches with regard to teacher effectiveness. According to the first approach, teachers learn to use certain teaching behaviors effectively and skillfully. In this connection, "effectively" means that the teacher can apply the learned teaching behaviors (teacher competence). Many courses have been developed in this field (for example microteaching). Within the second approach, "effectively" refers not to the teacher, but to the effects of his teaching. In the first approach, research examines whether the teacher can apply the acquired teaching behaviors, while in the second approach, it investigates the influence of the teaching behaviors on the student learning outcomes. It should be clear that the two approaches follow naturally from one another. When a researcher wants to study the influence of specific teaching behaviors on student learning outcomes (research into teacher effectiveness), he must be certain that the teacher can and does perform these behaviors (teacher competence). The results of research based on teacher effectiveness clearly demonstrate two points: first, relatively inexpensive courses can change fairly easily the behaviors of working teachers. Second, these training programs have a positive influence on student learning outcomes (Anderson, Evertson & Brophy, 1979; Crawford, Gage & Stayrook, 1978; Good & Grouws, 1979; Stallings, Needles & Stayrook, 1979). The desired changes in teaching behaviors do not demand a revision of the curriculum, nor a change in the organizational structure of the school, nor any newly trained staff, although we would in no way wish to imply that these changes are unimportant. The professionalization of teacher training and, consequently, the teaching profession is, at least we assume it is, naturally of great importance. In our opinion, teacher training programs should pay attention not only to teaching behaviors that increase the competence of teachers, but also to those that have a positive effect on student learning outcomes (cognitively as well as affectively).

The outline of this chapter is as follows. First, we will look generally at research into teaching behaviors. In emphasizing that teacher training programs should incorporate those teaching behaviors based on research results, we describe, as examples, four studies based explicitly on this concept. We describe two points where these studies differ from previous research into the effectiveness of teaching behaviors. Next, we stress the importance of systematically designed teacher training programs. As the design and realization of training programs should be not an art but a science, we describe a model for the systematic design of instruction in the case of training programs. By using this model, we describe the development of the training programs in the four model studies. We have recently completed our own correlational study (chapter 4). On the basis of our research results and those of other studies, we have designed a training program for mathematics teachers (Van der Sijde, 1985a). This program (the teaching script) was used to conduct a teacher competence (Van der Sijde, 1985a) and a teacher effectiveness study (Van der Sijde, 1986a). We will describe the design cycle followed in this study and elaborate on how these teaching behaviors can and should be incorporated into a regular in- service teacher training program.

1 STUDY OF TEACHING BEHAVIORS

As stated in the previous section, in the field of research into teaching behaviors we can distinguish between research into teacher competence and research into teacher effectiveness. Research into teacher competence explores the central question whether teachers can be trained in specific techniques or skills. An example to demonstrate this is the following hypothesis, that occupies a central position in Kieviet's research (1972): "In the mastery of specific teaching techniques prospective teachers will make better progress if they learn these techniques via microteaching than if they try the techniques out in a particular way in the real learning situation at school". Also the so-called mini-courses, by which we understand self-study packages based on microteaching, try to enlarge teacher competence (Allen & Ryan, 1969). The question whether changes in teaching behaviors have an influence on student learning outcomes has received verry little attention in the Netherlands in the past. Both microteaching and mini-courses are based mainly on the application of Skinnerian techniques to teaching skills: "small steps", "direct feedback", and "shaping by means of cues and prompts" (Heith, 1982). In short, these two types of teacher training are inspired mainly by theory and associated research, and are concerned with the process of change in the teacher.

The other approach, research into teacher effectiveness, investigates the influence of the teacher's instruction on student learning outcomes. As a rule, this type of research sets out from practice. What do teachers do in the classroom? What is the relationship between teaching behaviors and student learning outcomes? Gage (1984) calls this type of research the "observe-correlate-experiment" pattern and Rosenshine and Furst (1973, see also Dunkin & Biddle, 1974) the "descriptive- correlational- experimental loop". Teaching behaviors are observed, described, measured (quantitative description of teaching behavior), then correlated with learning outcomes, and finally "manipulated" in a training experiment. This last point touches on experimental research where the important correlational research variables are tested in strictly controlled situations. Gage (1985a) maintains that this pattern of effectiveness "does not depend on any particular set of educational values".

We cannot obtain an answer to the question as to which teaching behaviors are effective without conducting further research into it. We agree with Gilbert and Mosteller (1972) on the necessity for strictly controlled, ecologically valid field experiments.

1.1 Training Studies

As correlation coefficients between teaching behaviors and learning outcomes can give suggestions for reinforcing teaching behaviors which are considered desireable, such results can certainly serve as a source in developing a training program. The effectiveness of these recommendations must be tested experimentally before one can draw up definitive guidelines for teacher training. Considering the fact that use is made of the results of correlational research, it will be clear that the design of training programs is not in itself the goal of the series of research and design activities. We should regard the design of training programs as a component of the "observational-correlational-experimental loop", in which the designed, evaluated, and perhaps revised training program constitutes the experimental condition.

The training programs used in the studies occupy a central position in this chapter. According to Gage (1985b) up to 1985 there were thirteen known studies at the moment. A great number of these studies have only recently been completed (between 1982 and the present). However, four studies reached completion some time ago: Anderson, Evertson, and Brophy, 1979; Crawford, Gage, and Stayrook, 1978; Good and Grouws, 1979, and Stallings, Needles, and Stayrook, 1979. We will treat these four, the best known of the thirteen completed studies.

The four experimental studies differ in two important respects from research that has been done into teacher effectiveness in the past. First, these studies were based on the results of process-product correlational studies revealing relationships between classroom processes and student learning outcomes. Previous research frequently relied exclusively on one of the many psychological learning theories, which proved their value in a laboratory, but not in the classroom. In addition to these, rather a large number of studies have been carried out based on a particular philosophy of education. Such a philosophy does not usually evolve from a careful analysis of empirically established relations between educational processes and student learning outcomes, but often from anthropological, theological, or philosophy of life considerations. In the four studies mentioned above, however, it was first established which teaching behaviors correlated with learning outcomes. The data was then used for the design of a teacher training program. We do not doubt that training programs can also be based on a theoretical or philosophical notion; however, in our view, they must not be the only sources. Second, through systematic observation of the course of classroom events, the four studies were able to determine how far the teacher training program had had the desired effect on the teaching behaviors. That is to say that whether the teachers applied the behaviors they had been trained in, and if so, to what extent, were investigated. We can distinguish two phases in time in the four studies.

In the first phase, the teacher training program is regarded as the independent variable which is manipulated in the sense that the training is given to one group of teachers while the second group is withheld from the training in question. The dependent variable consists of teaching behaviors regarded as relevant and which can be measured by means of systematic observation in the classroom. In the second phase, the teaching behavior attained in the training program is the independent variable exhibited by the trained and untrained teachers. Student learning outcomes in a particular subject are the dependent variable. The results obtained in the first phase indicate how effectively the teacher training program influenced teaching behaviors (teacher competence). The results of the second phase reveal how far the teaching behaviors being incorporated in the training program really produce an improved effect on student outcomes (teacher effectiveness). An important aspect is that the researchers have explicitly attempted to do justice to "ecological validity"; the results are to a large degree representative of the course of events at school. The studies were conducted in regular classrooms, not in a laboratory. The period of teaching covered a substantial period of several months or even a whole school year, thus being in contrast with the period of no more than a few weeks or days covered in a great deal of prior research. Currently practicing teachers took part in the research. The manipulated teaching behaviors in the studies were realistic in so far that the teaching behaviors had already been observed in other teachers. Thus, from the beginning, some teachers demonstrated the desired behaviors. As a result, it was guaranteed that the independent variable (the teaching behaviors being incorporated in the training) did not happen to be strange and artificial to the teachers and students.

2 A MODEL FOR THE SYSTEMATIC DESIGN OF INSTRUCTION

Using a model for the systematic design of instruction, in this case training programs, we shall analyze the design and evaluation process of the training programs in the four mentioned studies.
Two stages must precede the establishment of a training program in a particular area. First, one should take an inventory of everything known about the area under consideration. In the training studies a training program for teachers was established which set out to improve the student learning outcomes (cognitive and/or affective). For this purpose, the researchers had to review the known research in this area. This is the pilot phase in which the problem is formulated and described in terms of the discrepancy between the actual and the desired situation. An analysis of the context, or at least the situation in which the problem arises, is also necessary for this. The definition of the problem, as well as the context analysis, provide the provisional critera that any potential solutions must meet. Based on the aim and the available knowledge in that particular area a proposal for a possible training program is made. This proposal is the design. We can outline the activities in the design phase as follows.
Starting from the definition of the problem, a solution is developed in this phase. To develop the solution, knowledge that is available through scientific sources as well as practical empirical knowledge is used. The proposals of alternative solutions and the comparison and evaluation of these alternatives, culminating in the choice of the most promising design are the characteristic activities in the design phase. The design phase must culminate in a working document, the guiding principle in the construction of the prototype training program. In our view, the design phase is important because the intentions and theoretical premises of the training program being developed become clearly visible in the design. The intentions and theoretical premises are of secondary importance in the training material, syllabi, computer- and videoprograms and suchlike. It should be possible to construct the training material as valid and scientifically sound, nevertheless the actual material should be suitable for the target group and for this group it is not necessary to be interested in the scientific and theoretical rationale underlying the training. Our colleagues who are also designing training programs, however, show an interest not only in the product, but also in its design and its rationale. Using the design, the components of the training are being developed. Obviously there will be interaction between the design and its realization. In the realization phase, certain problems will arise that lead to (a) modification(s) of the design. The final product of the realization phase will be a prototype training program. The prototype will be tested in practice. This is only possible by testing the training program on the target group and evaluating

it. In this phase, the proposed solution has to be tested and evaluated in the context it has been designed to determine whether the solution is satisfactory or improvements necessary. The evaluation is essentially formative. Two separate evaluations assess the training program within the terms of the training studies. The first evaluation point is whether the target group (practicing teachers) can apply the behaviors included in the training program. In other words, the question of whether teachers reveal more of the desired behaviors after completing the training program than they did before (teacher competence) needs to be investigated. Second, some training programs are developed with a specific aim, viz., improving the student learning outcomes by changing teaching behaviors (teacher effectiveness). Consequently, one must investigate whether the scores are indeed higher. Two experimental studies are necessary to answer these two questions. Only if both answers are affirmative can the training program (in this case) be offered to the target group. We call this the implementation phase.

3 ANALYSIS OF THE DESIGN PROCESS OF THE FIVE TRAINING STUDIES

By means of the different phases in the design process we investigate the extent to which there is a systematic approach in the design process of the training programs in the four model studies. The question towards the aim of the training program is stated in the pilot-study phase, along with the scientific sources to be consulted and the minimum conditions for the training design.
In the four studies, the answer to the question as to what should be achieved by a teacher training program has two aspects. First, the studies were intended to train teachers to improve their classroom arrangements. Concerning the teaching behaviors which had to be taught, the model studies included both low-inference and high-inference variables. Second, the studies intended to increase students' knowledge and to improve their attitude towards a particular school subject. Looking at the material used, we conclude that in all four studies in order to develop the training programs use was made of results of research carried out either by the researchers themselves or by others.
Anderson et al. (1979) used the results of process-product research carried out among reading instructors working with small groups of students in lower elementary school classes. On the basis of their own and other's findings, they formulated 22 guidelines for teaching. These were regarded as highly promising for the improvement of students' reading outcomes. Crawford et al. (1978) made use of results from four studies, namely Brophy and Evertson (1974), McDonald and Elias (1976), Soar (1973) and Stallings and Kaskowitz (1974). Like Anderson et al., Crawford et al. conducted their study in the lower classes of the

elementary school. For every significant correlation coefficient between teaching behavior and reading outcomes, the researchers specified operational descriptions of the observed behavioral variables, while at the same time stating the mean and standard deviations of the teaching behavior variables. This last point appeard to be useful in order to demonstrate that a high correlation coefficient does not mean that a particular form of teaching behavior must also frequently be present. In the general outline of Crawford et al. it appeared that some as specific, desirable behaviors were seldom exhibited by the more effective teachers, while the less effective teachers did not exhibit these behaviors at all. Using information gathered in a similar way, the researchers were able to compile a set of recommendations for teaching behaviors. Good and Grouws (1979) used results concerning arithmetic instruction in the fifth year of elementary school. They used their own research observations of more and less effective teachers. They combined teaching behavior profiles of the two distinct groups with research completed by others to develop material for a training program. Stallings et al. (1979) started with their own study and carried out research among a group of teachers trying to improve student reading outcomes in secondary education. They compiled a list of 50 variables in connection with student outcomes and their frequency of absence from school.

With the exception of Crawford and Gage (1977), none of the authors reports on the design of the training program. As stated above, we regard the design phase is important because the designers here take responsibility for the choice of variables to be included in the training program, as well as for their instructional theoretical and technological premisses. The solutions and/or products which have to be realized on the basis of the design are developed with respect to the target group. Consequently, the material should relate to the level and needs of the target group. In general, teachers with a few years practical experience do not feel a strong need for further explanations of instruction-theoretical and -technological premisses. They demand concrete, directly applicable information. The products should express this. Whenever the design is not explicitly reported, the research premisses can be lost along with the product's underlying considerations.

Next we will discuss the realization, the concrete products, and the procedures used during the training program and everything connected with it.

Anderson et al. (1979) report that two meetings were necessary to train the teachers. The teachers received a 33-page instruction manual. During the second meeting, which took place one week after the first, the participants took a multiple choice test to see whether they had understood the manual. The manual included guidelines for managing reading groups and indications on how to react to answers given by students. Crawford et al. (1978) constructed two

forms of training programs. In the first form, the participant teachers received five booklets of about ten pages each. A test included at the end of each booklet allowed the teachers to check up on whether they had understood the text. The other form also used these booklets, but the participants afterwards attended a two hour meeting with the designers of the training program. During these meetings the participants discussed, analyzed, and acted out the teaching behaviors in role-play. Good and Grouws' training program (1979) consisted of a 40-minute instruction unit given by one of the designers combined with a 45-page manual containing recommendations to be put into practice by the teachers. The manual described an exposition of parts of an arithmetic lesson. Two weeks later in another 40-minute meeting, the designers answered questions about the recommended teaching behaviors and discussed the difficulties encountered by the participants. Stallings et al. (1979) organized four two-hour training meetings over two and a half months. The topics of the training program, brought together in a manual, were: (a) recommendations for change in teaching behaviors setting out from the individual participant profiles established on the basis of observation; (b) suggestions for coping with students' behavioral problems; (c) guidelines for organizing and structuring lessons and materials; and (d) a summary of all the recommendations and a synthesis of the three previous meetings.

In the four model studies, the training programs were relatively short, and the content was compiled in manuals and explained during the meetings. In none of the studies an evaluation of the training programs in the sense of research into teacher competence did take place. By this we mean what we explained before: when a training program is designed and put into practice, an evaluative study must be carried out to investigate whether the teachers in the target group do or can apply that particular teaching behavior. When a training program is not first evaluated in the above sense one is entitled to ask oneself, what the value of such a study is and whether the conclusions drawn are valid. It might always be possible that the results that are found were achieved in another way than through the training program. From the research reports, we conclude that the extent to which the variables being practiced are implemented, is only investigated in the experiment in the four studies, although this was only after the question about teacher effectiveness had been answered. The authors of the four studies did not report any clearly distinguishable criteria, with the exception of Stallings, who used relative criteria. Nor did they take measures when it seemed that teachers in the experimental group did not manifest the relevant teaching behaviors in class (implementation). We conclude that in the four model studies too late the question concerning the implementation of teaching behaviors was posed. As we have indicated, this question should be asked before the evaluation phase of the systematic design model and not during the experimental phase.

To summarize, in each of the four studies, the design and evaluation phases are particulary striking because of the lack of information about them. All the studies briefly report on the pilot-study. They restrict themselves to references to the literature. Only Crawford and Gage (1977) report on a design for the training program. All of the studies include a description of the product, the training, and the training manual. However, none evaluate teacher competence or suggests revisions of the training programs following from the teacher competence study. Such an evaluation is important because it is precisely the aim of this phase to investigate whether and to what extent the training program is implemented. As all the teaching behaviors to be practiced are derived from results of process-product research, that is to say, all variables are observed teaching behaviors, cumulated with student learning outcomes, it is possible in principle to implement each individual behavior. However, it is important that the total collection of variables is implemented. A study of the measure of implementation, that is to say a study into teacher competence, must be carried out before the question of teacher effectiviness can be answered.

4 THE TRAINING PROGRAM DESIGN

With regard to the experimental part of the Dutch contribution to the CES an experimental training program for mathematics teachers was designed with the aim of increasing the effectiveness of classroom teaching. Van der Sijde (1984) reports on this design. For the design of the training use was made of three sources from which indications for effective teaching behaviors were drawn. The first source is the correlational study carried out in the framework of the CES (Tomic, 1985). As the second source we used results orginating from similar research. The last source is the theory and research in the field of instruction.

Teaching behaviors from the correlational study

From our own research, it appears that there are correlations between a number of teaching behaviors and student learning outcomes in the cognitive domain: (a) the teacher asks a group or an individual student a high-level question within a large group; (b) the teacher rephrases a question, followed by student, in turn followed by an interaction other than redirecting or rephrasing the question, error, punishment, or the teacher answering the question him or herself; (c) the teacher asks a large or small group a question, followed by a silence of at least five seconds; (d) the teacher redirects a question to another student,

who answers the question correctly; (e) the teacher gives positive feedback to a student's answer.

In the affective domain, the following behaviors are correlated with student learning outcomes: (f) all forms of explanation given by the teacher; (g) the teacher rephrases a question, followed by a correct student answer; (h) the teacher gives positive feedback to a student's answer; (i) students listen to the teacher giving an explanation; (j) students take part in the discussion of previously covered work or with material dealt with previously, guided by the teacher.

Along with these general results, use is also made of specific characteristics of the target group. For our study, we designed a training program for mathematics teachers in the eighth grade using the Sigma-textbook. From Krammer's study (1984) it appears that textbooks elicit particular teaching behaviors. For this reason, we have based our study on the specific process-product correlations of the Sigma textbook.

Teaching behaviors originating from similar research

The research carried out by Good and Grouws (1979) also concerned teaching behaviors in mathematics instruction. Good and Grouws recommend what they call "key instructional behaviors". Based on their positive experience with such instruction behaviors, we have designed a similar sequence that should not be artificial but have to connect with the practice of mathematics instruction. In the appendix we compare Good and Grouws' sequence of instruction with our own. With regard to the further design of the training program, we profited a great deal from an article by Crawford and Gage (1977) dealing with their design.

Teaching behaviors derived from theory and research in the field of instruction

We have mentioned the variable "all forms of explanation given by the teacher". As this process variable is too global, we have looked for references to "explanation" in the literature and research which has already been carried out (see Boekaerts, 1978; Wardenaar and Willems, 1984). From the literature, it appears that summarizing (Hartley & Trueman, 1982) and working with examples (Merrill & Boutwell, 1973) promotes the retention of information. For this reason, we incorporated "summarizing" and "working with examples" into our teacher training program. From the correlational study that has been carried out it appears that along with the number of variables that refer to instruction, classroom management is also of great importance (Tomic, 1985). Management activities such as those studied in the correlational phase of the study leave little room to derive concrete teaching behaviors. Classroom management ("keeping control") is naturally of great importance. The great interest in management is also evident from publications in that area, for example Vonk (1984), Creton and Wubbels (1985), and Van der Sijde (1985b). We based recommendations for classroom management in our teacher traininig program on the work of Kounin (1970).

Before going into detail with respect to the specific recommendations that form the core of the training, we will first present a thorough theoretical framework as a reference for these recommendations. We start out from the point of view that teaching is a complex cognitive skill, amendable in a manner similar to other cognitive skills. This view led reseachers to investigate teaching expertise, but the problem with these studies is that it is hard to indicate the components of teaching expertise. We will treat this problem first.

5 KNOWLEDGE SYSTEMS IN TEACHING

Leinhardt and Smith (1985) assume that expertise involved in the cognitive aspects of teaching has two areas of knowledge: lesson structure knowledge and subject matter knowledge. The first is the knowledge required to construct and conduct a lesson, and the second is the knowledge of the content to be taught. These areas of knowledge interact. Berliner (1986) also distinguished knowledge of organization and management of classrooms. In the remainder of this chapter we will first discuss lesson structure knowledge, involving the skills needed to plan and run a lesson smoothly, and pass easily from one segment to another, and explain the material clearly (Leinhardt & Smith, 1985). Second, we will treat subject matter knowledge, and, finally classroom management knowledge.

5.1 Lesson Structure Knowledge

The main feature of the skilled teachers's knowledge structure is a set of schemata for teaching behavior, including structures more or less general - some concerning schemata for quite global behavior, such as checking homework, and some concerning smaller units of acitivity, such as distributing a paper to the class (Leinhardt & Greeno, 1986). Teaching thus, consists of a number of behaviors and each activity can be regarded as a set of schemata. We can also view teaching as based on a particular set of schemata, namely a script. Leinhardt and Greeno hypothesize that conducting a lesson follows an operational plan, which they call an agenda. When teaching is regarded as a script- performance, the actual teaching is based on entering the teaching script.

Entering a script means that one knows what to do, from moment to moment: a script points out what is appropriate for a particular situation (Schank, 1982). For example, someone is behaving the script "the patient during a dental check-up" (Abelson, 1981). This means that he must know what it is to be a dental patient, and that he has to have a reason to visit a dentist (painful gum, for example), and he has to decide to visit the dentist.
Teaching can also be considered scripted behavior. A teacher should have a stable representation (a cognitive image) of what teaching is and of his particular TEACHING script, which guides his interactive teaching (Shavelson & Stern, 1981). Such teaching scripts become routine. Once begun, they are typically completed. The teacher learns the script, containing some or all of the elements or scenes in his teacher training course(s). A stable representation does not mean that one has the correct, complete or ideal script. For example, Schank (1982) relates the story of his daughter (age 4). When he told her that he would buy a new car, she asked him if he was also going to buy a new key chain. "Why", he asked her? "Because", she said, "last time you bought a new car, you also bought a new key chain". Both events, buying a car and buying a key chain, were, for her, part of the same script. Through the later experience she must have learned that the two behaviors are unconnected and do not belong to the same script.
Teaching has different scenes, or as Leinhardt and Greeno (1986) prefer to call them, segments. These scenes are: (1) Homework check; (2) Presentation; (3) Monitored practice; (4) Guided practice; (5) Tutoring.
Each scene has its own specific activities. Starting with the general characteristics, which Leinhardt and Greeno assign each of the "segments", we will elaborate further in the next sections.

5.1.1 Presentation

Presentation is the most central scene in the teaching script. When a teacher enters this part of the script, he teaches new things, explains. In this process the teacher facilitates the meaningful acquisition of an idea by the students (Good, Grouws & Ebmeier, 1983). Besides, presentation facilitates the meaningful acquisition maintenance and transfer of newly learned skills (Kameenui, Carnine, Darch & Stein, 1986). Different authors have different labels for this activity: presentation (Leinhardt & Greeno, 1986; Rosenshine & Stevens, 1986); development (Good, Grouws & Ebmeier, 1983; Kameenui et al., 1986); instructional explanation (Duffy, Roehler, Meloth & Vavrus, 1986).
The different labels are synonymous. Based on qualitative analysis of interactions in the classroom, Duffy et al. (1986) describe four characteristics of presen-

tation. The first characteristic is response information-giving. Effective explanations are information giving experiences which supply students with the information they need to build or modify a schema. Developing awareness is a second characteristic. Students must become aware of the use and importance of what was presented during the lesson. Awareness takes four important forms: (1) metacognitive awareness: students should come to the right conclusion or answer, and know how they reached it; (2) awareness of what is being taught: students must learn when and how rules and strategies can be used; (3) awareness as an attribute of expert performance; (4) awareness as an information interpreter: a teacher emphasizes when and how to use the lesson content, students are likely to make the intended interpretation. The nature of information is another characteristic. A teacher who presents information should pay attention to (1) the kind of information (effective teachers use three types of information: declarative, conditional, and procedural knowledge. Declarative knowledge is knowledge about the task, its characteristics and how it is structured. Conditional or situational knowledge involves when and why a particular strategy is used. Procedural knowledge concerns how to apply a successful strategy); (2) the clarity of the information (the information a teacher gives must be precise and explicit); (3) the accuracy of the information (the teacher has to communicate correctly); (4) the usefulness of the information (information must not be presented without context). The last characteristic is "hook". Effective teachers present information gradually and organize it sequentially, and they embed assistance in their interactions with students, designed to help them restructure explanations in the intended way. Duffy et al. (1986) conclude that teachers ("effective explainers") explicitly describe at the early stages of a lesson the what, when, and how of the lesson. During a lesson, they provide the students with reminders, reviews, and summaries.

Putting together ideas derived from experimental studies (mainly training studies and primarily the work of Good, Grouws and Ebmeier, 1983), Rosenshine and Stevens (1986, see also Rosenshine 1986) produced a list of behavior during presentation depicted in Table 5.1. This table compares the cognitive approach of presentation (Duffy et al., 1986) with the quantitative approach of presentation (Rosenshine & Stevens, 1986).

Table 5.1 reveals that researches on teaching starting from two different approaches converge in their implications for teaching practice.

Table 5.1 Comparison between presentation activities (a qualitative approach (Duffy et al., 1986) and a quantitative approach (Rosenshine & Stevens, 1986)

the qualitative approach (Duffy et al.)	the quantitative approach (Rosenshine & Stevens)
-make explicit statements at stages of the lesson about what, when, and how to learn	-provide short statement of objectives
-during lessons provide structure reminders, reviews, and summaries	-provide overview and
-present information gradually and organize it sequentially	-highlight main points -proceed in small steps but in rapid pace
-present useful information, information must not be presented without context	-provide sufficient illustrations and concrete examples
-give effective explanations (information-giving experiences supply students with the information they need to build or modify a schema)	-provide demonstrations and models -when necessary, give detailed and redundant instructions and examples

5.1.2 Monitored practice

In monitored practice the students do seatwork (students work on presented problems at desks or at the blackboard) while the teacher moves about the room, occasionally tutoring the students while they work. According to Doyle (1986) little is known about organizing ("orchestrating") seatwork, because researchers of classroom interaction have a dominant interest in other topics. Management studies show that the effective teacher monitors the class thoroughly and, inspects individual papers frequently (Evertson & Emmer, 1982; Emmer & Evertson, 1981). During monitored practice, the teacher is likely to limit whole-class comments to reprimanding disruptive behavior (McDermott, 1976). Further, Atwood (1983) found that task complexity is negatively associated with engagement.

5.1.3 Guided practice

During guided practice, the teacher presents the students with a problem. It usually starts as seatwork all students trying to solve the problem individually. The problems in this lesson phase are more complex than during monitored practice; not only are newly presented knowledge (concept and principles) and procedures required to solve the problem, but also knowledge acquired during previous lessons, analysis of the problem, and synthesis of knowledge. The problems, in general, are a little too difficult for the students to solve on their own, so after a few attempts, the teacher guides the students to the correct solution of the problem. The problem is written on the blackboard (or overhead sheet) and divided into subproblems. The teacher involves all students in this process by asking, probing and redirecting questions. He provides the students with feedback on the answers to the (sub)problems and the procedures they use.

5.1.4 Tutorial

A tutorial is a form (method) of instruction where a student or a small group of students receives individualized instruction. In the activity the teaching script, it is remedial and it focusses on students who, during other lesson phases (e.g. homework or presentation), encountered such individual problems (not of interest to the whole class) that supplemental instruction only for students (or groups of students) is more effective.

5.1.5 Homework

Homework should take two forms in a lesson. At the start of a lesson, homework has to be checked; at the end, students must have the opportunity to start their homework for the next lesson.
Homework has a substantial effect on student achievement. A summary by Walberg, Paschal and Weinstein (1984) shows that an effectsize of graded homework is 0.80 and of assigned homework, 0.30. Why is homework effective? Homework increases time-on-task and as Walberg et al. (1984) state, time may not be the only ingredient of learning, but without it, little can be learned. Homework should not be misused, and Good et al. (1983) stress that perhaps the most devastating misuse of homework is when students are assigned problems without inadequate background supplied in class.
Homework must be checked. By checking, the students receive feedback and the teacher gains information about the students' progress and about his teaching effectivity (what concepts should be retaught?).

5.1.6 The sequence of scenes

The ideal lesson must contain all five scenes. Each scene addresses a unique aspect of the teaching-learning process. Because teachers differ (Good, Biddle

& Brophy, 1975), variants of the script arise in different sequences. For example, some teachers like to start a lesson by explaining new concepts and then proceed to correct homework, so the students who did not encounter problems can start their new homework assignments. In the script, the first two scenes are thus reversed. In another, the scenes on seatwork are sometimes considered equivalent, although with some major differences. Galambos and Rips (1982) presented a slightly different concept of the script-notion. Instead of scripts, like the restaurant-script below, they refer to such notions as "routines" and to their component parts (scenes) as "episodes". To indicate their different view, they discuss the routine "changing a flat". This routine can be represented in memory according to sequence (see Figure 5.1).

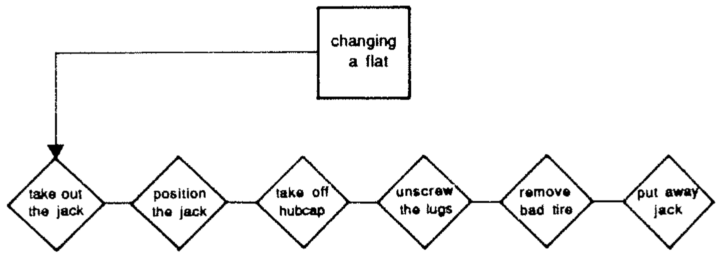

Figure 5.1 Sequential representation of a routine (source: Galambos & Rips), 1982)

Sequence means what follows what. However, we can represent a routine in memory a different way, namely, according to the centrality of the episodes (N.B. Rummelhart's (1977) model of story structure is such a centrality-oriented theory). The more important episodes are more centrally represented (see Figure 5.2). In the routine "changing a flat" the episode "take out the jack" precede "unscrew the lugs" (sequence), but "unscrew the lugs" is more central than "take out the jack". Script theory designates certain episodes as MAINCONS (main conceptualizations) if they are vital for the fulfillment of the script action, and these MAINCON episodes are what Galambos and Rips (1982) call central episodes.

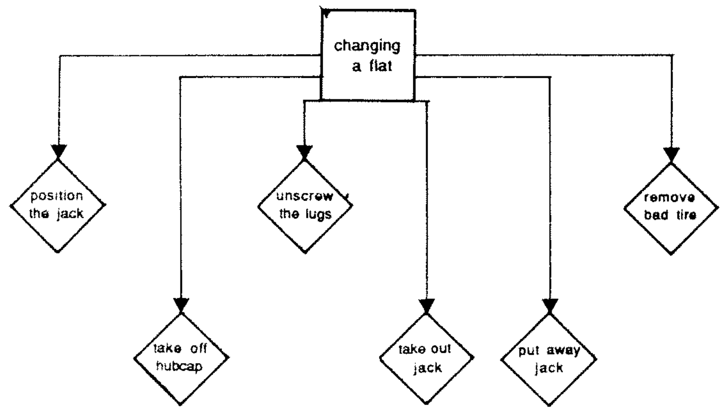

Figure 5.2 Centrality representation of a routine (source: Galambos & Rips, 1982)

Teaching can be considered a script (analogous to the ideas of Schank & Abelson, 1977) or as a routine (analogous to the ideas of Galambos & Rips, 1982). However, the teaching script is very different. For example, the restaurant script. Going to a restaurant (entering) and leaving (exiting) without having ordered and eaten means the restaurant script was not entered. The most central episode of the restaurant script (or routine), eating, was not fulfilled. The teaching script (or routine) is very different. The teaching script is entered even if the teacher only completes the presentation scene, or checks the homework and assigns new homework. The script is also entered if the teacher completes the monitored and guided practice scenes, but not if the only scene completed is tutoring. A centrality represented teaching routine is depicted in Figure 5.3.
Completing the most central episodes of the routine, means, in the case of teaching, completing the routine (the script has been entered). In the restaurant-script and in the changing-a-flat-script, the scenes are interdependent, while in the teaching script some scenes are so independent that completing that particular scene means completing the script.
Because teachers do not follow the same procedures, they execute the script differently. These differences become observable in different orderings of the script scenes, but also in the omission of scenes.

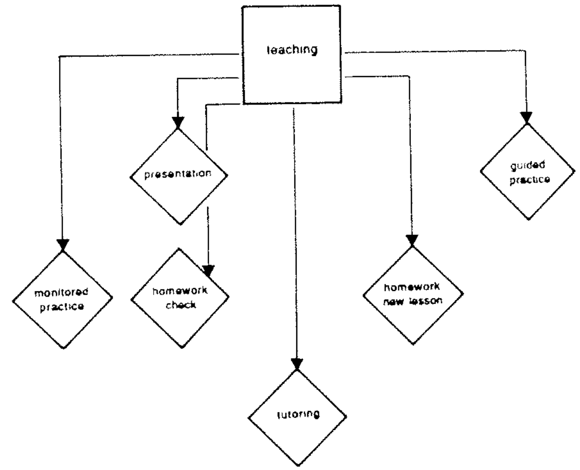

Figure 5.3 Centrality representation of the teaching script

5.2 Subject Matter Knowledge

Lesson structure knowledge is supported by subject matter knowledge. "Subject matter knowledge", as Leinhardt and Smith (1985) write, "includes concepts, algoritmic operations, the connections among different algoritmic procedures, the subject of the number system being drawn upon, the understanding of classes of students errors, and curriculum presentation. Subject matter knowledge supports lesson structure and acts as a resource in the selection of explanations, and demonstrations". Subject matter knowledge also depends on the textbook: the information to be learned is the basis for designing instructional strategies (Tennyson & Cocchiarella, 1986). The research reported in this book confines itself to mathematics in the second grade of the Dutch secondary school.

5.2.1 Second grade mathematics

As part of the Second Mathematics Study (an IEA-project) in the Netherlands, Pelgrum, Eggen, and Plomp (1982) sampled information on the topics covered in the second grade of secondary schools. They concluded from their study that 14 topics (see Figure 5.4) were covered.

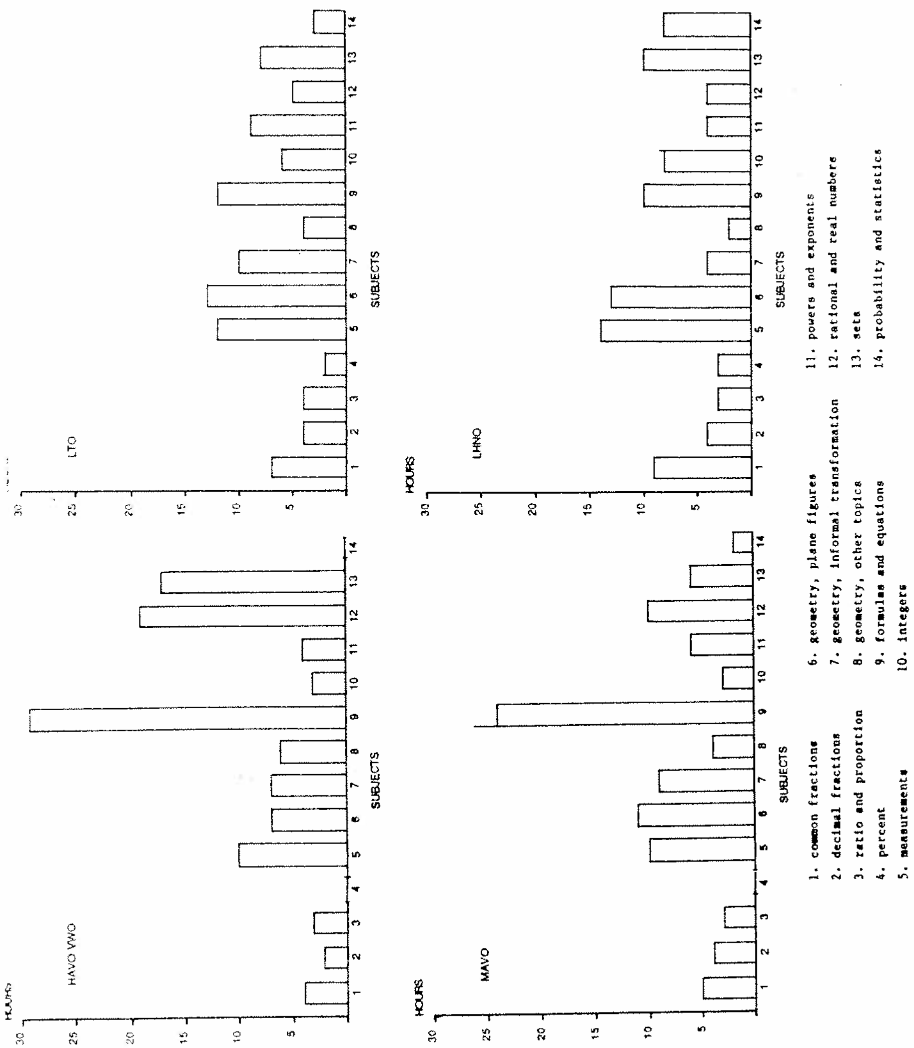

Figure 5.4 Relative allocation of hours for each mathematics subject in each school type

Figure 5.4 depicts the relative proportion of time devoted to each of those subjects. Apparently the Mavo, Havo and Vwo relatively pay more attention to formulas and equations, while the Lbo (which includes Lto and Lhno) distributes time over the different subjects. This certain uniformity in curricula and subject matter depends on factors such as examination requirements and the textbooks used, mainly Moderne Wiskunde (Modern Mathematics), Getal en Ruimte (Number and Space) and Sigma. Krammer's research (1984) shows that the textbook also influences teaching behavior, for instance the use of examples during presentation.

Recently, a new curriculum reform (HEWET: Herverkaveling wiskunde een en twee - Reparcelling math one and two) has been implemented. Although it primarily affects the subject matter in the higher grades, it also has an impact on second grade subject matter. The HEWET program discusses the role mathematics can play in society; mathematics becomes more concrete and less abstract. This "new" approach also appears in the textbooks for the second grade.

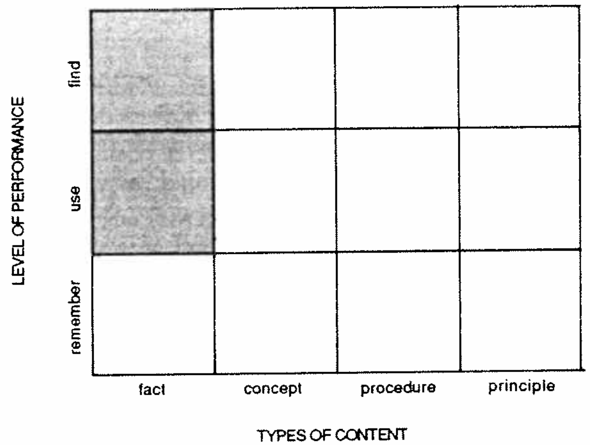

Figure 5.5 Performance-content matrix (source: Merrill, 1983)

5.2.2 A characterization of the mathematics content based on Merrill's Component Display Theory

Mathematics in the second grade Havo and Vwo can be characterized using Merrill's performance-content-matrix. The performance- content matrix (Merill, 1983) forms the core of Merrill's Component Display Theory (CDT), comining different levels of (student) performance and types of content (Figure .5).

'he level of performance means the performance a student must deliver; member means recognition or reproduction of some item of information; and se means the utilization of information in a specific case; and find means acvity requiring a student to deduce or invent a new abstraction. The "type of ontent" refers to the type of information to be learned. This might be a *fact* (an olated piece of information), a *concept* (a group of objects, events, or symbols hat all share common characteristics), a *procedure* (an ordered sequence of eps required to accomplish a goal), or a *principle* (an explanation or predicon of why things happen). Every combination of content and performance is ossible (except use-fact and find-fact).

he building blocks of mathematics are concepts and principles and means to pply them, but performance generally consists of "use" and "find". Examples: Solve in R: 5x + 1 2x - 5", and, "The length of a diagonal of a square is 1, callate the area".

3 Classroom Management

ne of the research-questions in classroom management compared the difrences between a good and a less good classroom manager. A good classroom anager has a successful classroom and successful classrooms are defined as hose having a prevalence of work involvement and a low amount of misehavior in learning settings" (Kounin, 1970, p. iv). The results of the comarison were rather surprising, because both good classroom managers and less ood classroom managers used virtually the same curative techniques to restore try to restore the order. Nevertheless, good and less good managers differed the prevention of classroom disorders. In the next section we will discuss techques to prevent classroom disorders observed in the behavior of good classom managers.

uke (1979) describes classroom management as the provisions and procedures ecessary to create and maintain a situation where learning and teaching can ke place. The label "classroom management" stands for a variety of teacher anagement behavior: management of behavior; management of instructional sks; management of thinking process (Soar & Soar, 1979). The difference beveen management of instructional tasks and thinking process is a gradual one nd this study will treat the two. Management of behavior concerns those beavioral aspects in the classroom that do not relate to the task to be performed. vertson and Anderson (1979) call this off-task behavior and further distinguish

sanctioned (off-task behavior, but no misbehavior) and non-sanctioned behavior (not paying attention, talking, etc.).

A teacher performs classroom management, particularly management of behavior, to keep order in the classroom. Cohen, Intilli, and Robbino (1979) describe order as "the situation where there is a clear set of expectations for all classroom members, where people can anticipate how others will behave, where people feel that it is right and proper for everyone to conform to these expectations, and where there is a high degree of conformity to the expectation" (p. 118).

Order is a condition for instruction. Without order in the classroom, instruction (teaching) is not possible. Every classroom should have a set of clear agreements on what is allowed (sanctioned off-task behavior) and not allowed. Disturbances must be prevented to avoid disciplinary measures. Management of instructional tasks is the manner instructional tasks are selected and presented by the teacher. Presentations of problems and tasks must be sufficiently varied and challenging to keep the students interested and motivated. Management of instructional tasks requires the teacher to create conditions students will and can perform tasks and support students who encounter problems. We distinguish between on-task behavior (listening to the teacher, reading, and writing) and procedural behavior (task preparing: execute a procedure or a routine which is not on-task behavior, but which behavior is expected and desired by the teacher, e.g. distributing textbooks).

The transition between behaviors, presentation, monitored practice must be as smooth as possible. Transitions generally contain more off-task behavior than other periods of a lesson (Arlin, 1979). Classroom management in general, and management of instructional tasks in particular, are bound to enhance on-task student behavior.

5.3.1 Management techniques according to Kounin

Kounin identifies six important management techniques in his study of classroom management behavior. These management techniques are: withitness ("eyes in the back"); overlapping (two things at the same time); smoothness and jerkiness (the course of the lesson); momentum and slowdowns (continuous pace); group focus (attention on the group); anti-satiation (student surfeit).

These techniques can be divided into three groups, namely withitness and overlapping, lesson pace (smoothness and jerkiness, and, momentum and slowdowns), and group dynamic (group focus and anti-satiation).

5.3.1.1 Withitness and overlapping

Kounin (1970) considers withitness and overlapping a teacher's two most important management techniques. Withitness can be defined as follows: a teacher communicates with his students through his (overt) behavior; he shows that he knows what the students are doing and what goes on in the classroom. He has the proverbial eyes in the back. Overlapping is a teacher's doing two or more things at the same time; he distributes his attention over different things.

Two situations may occur. In the first, a teacher is engaged in classroom instruction and management of behavior. For instance, a teacher gives instruction while a number of students do not pay any attention (they are showing off-task behavior), and he rebukes them for their behavior. In the second, the teacher instructs and manages instructional tasks. For example, the teacher is explaining a couple of things to a group of students while one student (not belonging to that group) asks the teacher for advice, which the teacher then gives.

The teacher shows overlapping and withitness in only the first situation; in the second situation he only shows overlapping. Withitness is "overt" behavior, while overlapping is "deducible" behavior. Withitness can also occur without overlapping. In this case a situation is "open": a teacher, for example, sits behind his desk (or walks through the classroom), while students do their work (e.g., monitored practice). When, in such a situation, students show off-task behavior and the teacher reprimands them, he is showing withitness (overlapping is out of the question; the teacher only performs management behavior).

In Kounin's research (1970) into good classroom management, it appeared that withitness explained more variance than overlapping. This is because withitness is directly observable and overlapping is not directly observable. The correlation coefficient reported between withitness and overlapping is $r = .60$.

Withitness is the visible behavior showing that the teacher knows what goes on in the classroom. Of course, this behavior is only effective when he calls the right student(s) to order at the right moment. When he fails, he shows that he is not suffiently withit.

Irving and Martin (1982) call withitness the "confusing variable", because Kounin's research does not show how withitness should be measured. This becomes clearer from their (partial) replication: they found significantly different results concerning the relation between withitness and decreasing deviant behavior (z-score of the difference: $z = 2.74$, $p < .01$) and increasing task involvement (z-score of the difference: $z = 2.94$, $p < .01$). Irving and Martin (1982) ascribe their deviant results to differences in design and to the "confusing variable" withitness.

5.3.1.2 Movement management

A teacher has a lot to do in the classroom and his behavior must be as smooth as possible. The progress of a lesson must not be slowed (momentum) and he must avoid sudden interruption (smoothness). In more recent publications, Kounin (Kounin & Doyle, 1973; Kounin & Gump, 1974) speaks about lessons as signal systems. A lesson should have a continuous signal that should not be interrupted. The signal comes from the teacher as well as from the students and the textbook.

Kounin's research shows a high correlation between momentum and smoothness ($r = .75$). When the speed of a lesson is irregular, the students engage in more off-task behavior.

5.3.1.3 Group focus

Teachers only rarely instruct a single student. As a rule, they teach a class of students or a group of students (Goodlad, 1984; Cuban, 1984). Because of this, group focus is of major importance. Kounin made three distinctions: format (participation of students. Bloom (1976) calls it an instructional variable); involvement (the attempts a teacher makes to involve students who show off-task behavior); and student responsibility (the degree the teacher holds the students responsible for the execution of their tasks during the lesson).

5.3.1.4 Variation

Although a lesson must exhibit a continuous stream of behavior, this does not preclude variation. On the contrary, a lesson should contain different activities to prevent student boredom and off-task behavior.

5.3.1.5 Types of management behavior

In the beginning of this chapter, we distinguished between management of behavior and of instructional tasks. The techniques concerning withitness and overlapping, involving the group, and student responsibility rank under management of behavior, while momentum, smoothness, format, and variation are part of instructional task management.

5.3.2 Towards a cognitive model for classroom management of behavior

"Efforts need to be devoted to the development of cognitive models in classroom management, that is, frameworks for representing the understanding teachers have of management process and classroom order" (Doyle, 1986).

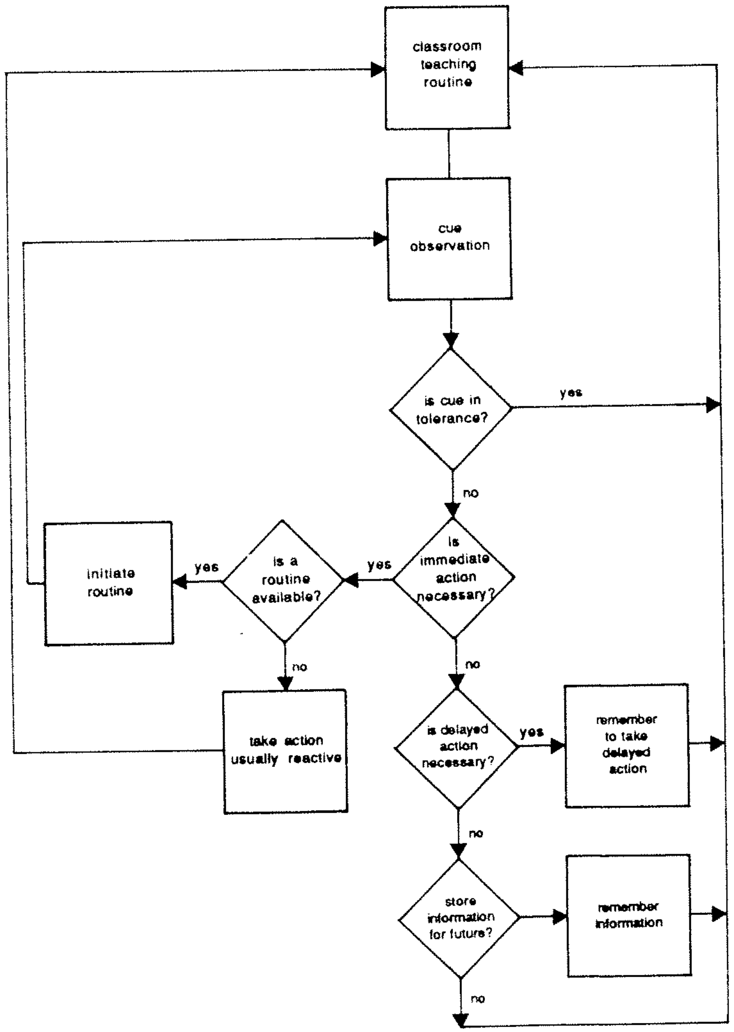

Figure 5.6 Model of teachers' decision making during interactive teaching (source: Shavelson & Stern, 1981)

5.3.2.1 The model of Shavelson and Stern

Shavelson and Stern (1981) present a model (see Figure 5.6) for interactive teaching where the concept "routine" plays a central role. It can be described as a fixed procedure or program whose main function is to control and coordinate a specific sequence of behavior. Such a sequence is a familiar non-problematic and efficient way of acting in identical situations (Hol, 1983). In the classroom the teacher performs the classroom teaching routine and he receives signals from the class ("cue observation"). Clark and Peterson (1986) mention as the greatest merit of this model that interactive teaching occurs when the classroom teaching routine is interrupted. Based on the strength of the signal, the teacher decides whether it is tolerable or not. If not, he decides whether action seems appropriate, and if so, whether he has a routine for it. If the teacher has a routine to restore the signal to within limits, he will execute it. If not, he has to act spontaneously. Some cases do not call for immediate action, and a delayed action seems appropriate. From a Kouninian point of view, a number of remarks can be made. First, we need not distinguish between an immediate and a delayed teacher action, because an action should be carried out at the right moment and directed towards the right student(s). The "crucial" moment of decision depends on the signal ("cue observation"). No (re)action taken at the right moment can lead to greater classroom disturbances, that cannot be corrected by simple "routines". Another remark concerns the concept of "routine". Kounin does not use this concept; he describes a number of techniques useful for a teacher. A teacher action in Shavelson and Stern's model is either a routine or a spontaneous action. The techniques of Kounin are practical and useful in the classroom. Teachers can apply a new technique before it becomes routinized. This new technique is neither a spontaneous action (it is a well-considered selection among the teacher's possibilities) nor a routine. Clark and Peterson (1986) mention this as a shortcoming of the model.

5.3.2.2 Towards a model based on Kounin's theory

A model design based on Kounin, uses his most important variables/techniques: withitness and overlapping. The relation between those two variables can be illustrated with the following three examples:

Example 1: The students of grade 8 are busy doing some of their homework assignments in the classroom. The teacher is helping Caroline to solve problem 2. Suddenly Mark starts talking to Paul. The teacher notices this and asks the two boys to continue their homework and not to disturb the rest of the class. Meanwhile, Caroline has continued solving the problem. The teacher helps her again but keeps watching the boys.

Example 2: The students of grade 4 are doing a linguistic exercise and the teacher is giving a group of four students some extra instruction. At the time, Margaret raises her hand and asks for help. The teacher responds very quickly while keeping an eye on the group he was instructing. After helping Margaret, he continues working with the group of four.

Example 3: The students of grade 9 are reading a text and answering questions. The teacher is standing at the back of the classroom. When Peter throws a pellet, the teacher asks him to pick it up and continue working.

In example 1, the teacher is showing overlapping as well as withitness. He is busy with one student, but notices at the same time that some students are doing something else. The teacher judges this signal as disturbing and makes a remark to those students.
Example 2 exhibits only overlapping. The teacher attends to a signal from the class very briefly and afterwards continues what he was doing.
Withitness is demonstrated in example 3; the teacher monitors the students during seatwork, receives a disturbing signal and undertakes action.
These examples do not always clearly illustrate, whether both withitness and overlapping are present. During overlapping a teacher receives signals from the students. He has to decide whether he should attend to these signals or not. Attending to the signals means showing observable behavior (observable to the students). The observable behavior results from a cognitive process. During his main activity (instruction and/or management of instructional tasks) the teacher continuously judges signals coming from the students. In other words, during instruction when the main signal comes from the teacher (or from the textbook in case of management of instructional tasks), the students signal is continuously judged and compared with the intended signal (from the teacher or textbook) for possible interference. When the two signals do not interfere (do not disturbe), there is no need for the teacher to undertake any action. But when they do interfere, when the signal from the students is not within tolerance, then the teacher has to undertake an action (see Figure 5.7).
Undertaking an action usually means showing withitness. As example 2 illustrates, not all teacher actions are intended to restore order, but if such a student signal remains unnoticed, students will possibly start showing off-task behavior and disturbe other students. A teacher not only judges if the students' behavior do disturb the ongoing activities (interfering student behavior/student signals), but also must be on his guard against possible disturbances which might interrupt the ongoing activities (interference of the teacher's or textbook's signal with the students' signal).

This model is rather global and will be further elaborated in the next section.

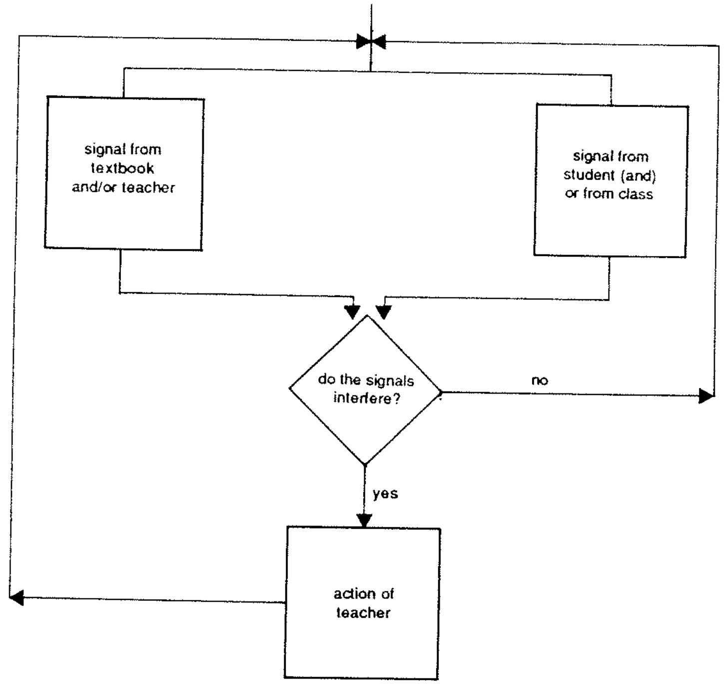

Figure 5.7 A global model based on Kounin (1970)

Elaboration of the global model Two of the three previous examples described situations with overlapping behavior. The first example describes overlapping of instruction and management of behavior; the second describes overlapping of instruction and management of instructional tasks.
In overlapping of instruction and management of behavior, the teacher performs both behaviors at the same time. During classroom teaching, he continuously monitors the students. Signals coming from the students that do not disturbe the teacher activity are no motive to call the students to order. A motive exists when the signals threaten to disturb order. In that case, the teacher should act: interrupt teaching and carry out an appropriate action to restore a workable teaching situation. Figure 5.8 elaborates the global model.

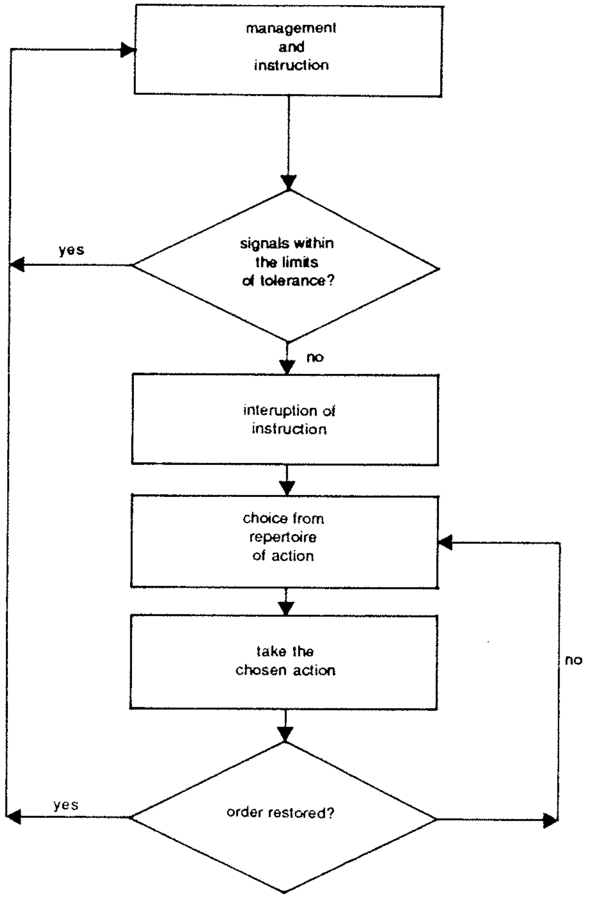

Figure 5.8 Overlapping of management and instruction

Overlapping means overlapping of instruction and management of students' behavior (teaching), together with the constant evaluation of signals coming from the class (see Figure 5.8). When the teacher undertakes an action (interrupts instruction, selects and executes an action), he shows he is withit. Withitness in this situation is the visible result of overlapping.
This process describes in Kouninian terms a chain of events which Shavelson and Stern (1981) call interactive teaching.
Example 2 shows overlapping of instruction and management of instructional tasks. There is overlapping, because during the instruction (to a part of the class),

the teacher is monitoring the other students. He receives a signal, not a disturbing one, from a distinct student. This student tries to attract the teacher's attention because she has a problem. The teacher does not have to show withitness because the two behaviors can alternate.
Example 3 shows overlapping of management of instructional tasks and management of behavior. To separate these two management tasks and call this "overlapping" would be artificial. The teacher monitors the classroom, and depending on the nature of the signal he shows management-of-instructional-tasks or management-of-behavior. In this situation a teacher shows that he is withit (see Figure 5.9). Shavelson and Stern would not call this situation interactive teaching.

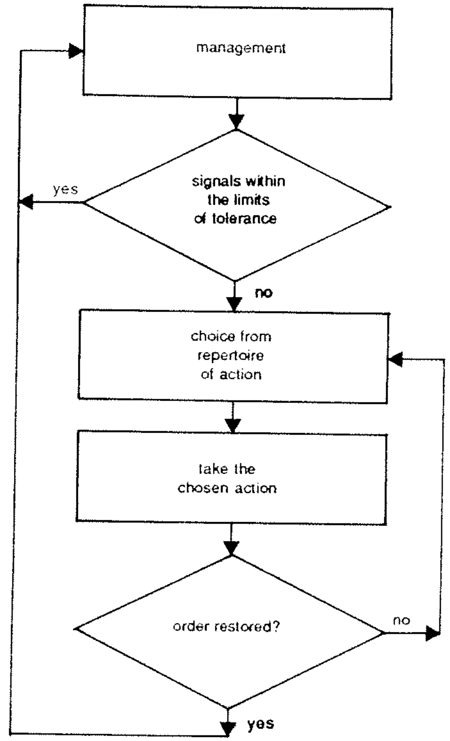

Figure 5.9 Overlapping of management of behavior and of instructional tasks

5.4 Implications for Teacher Training

The previous sections present models for instruction and management, that can be combined to one model. In Figure 5.10 we present a model for classroom teaching. During teaching, there is constant "input" from the class, and students. When a teacher starts a lesson, he must decide whether the signals from the class (the "input") are within limits. If they are, he can start his instruction; if not, he must select and carry out an action in order to restore order (to reduce the input signals of the class). If order is restored, he can start instruction. During instruction, a teacher constantly checks whether or not he can continue instructing. If at some point the signals of the class are disturbing, the teacher interrupts instruction (and remembers where he stopped), selects and carries out an action, and checks if he can continue (where he left off).

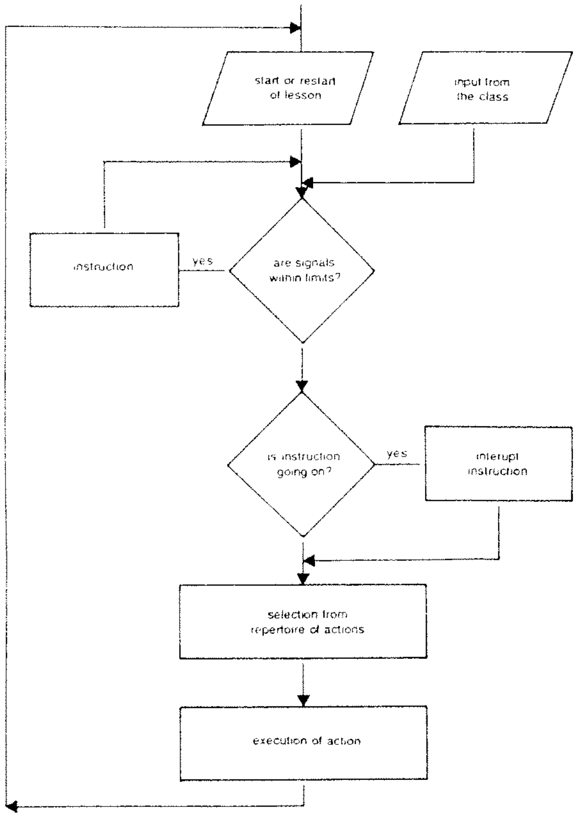

Figure 5.10 A model of classroom teaching

The instructional model, represented in Figure 5.10 by the box "instruction", contains the five scenes "homework check", "presentation", "monitored practice", "guided practice", and "tutoring".
There are several alternatives for a teacher training. Starting from the model of classroom teaching depicted in Figure 5.10, we have three points of departure: instruction, selection of an action, and execution of an action. The latter two are usually combined into management. For instruction we have two approaches. The first is developing a teacher training in order to enhance the implementation of a particular (part of a) curriculum (see, for example, Snippe & Harskamp, 1986). The second is developing a teacher training in order to enhance teaching effectiveness. In the latter, general subject-independent rules are incorporated in the training.
For management we also have two approaches. The first stresses extending the repertoire of actions, and the second also incorporates opportunities for practice in reduced learning environments (for example, microteaching).
Independent of the chosen approach the training should contain rules for management and instruction. Using those rules, teachers can design their lessons; teachers are instructional designers for their own lessons. Instructional design is concerned with understanding, applying, and improving methods of instruction. It is the process of deciding what methods are best for bringing about the desired changes in student knowledge and skills (Reigeluth, 1983). The result of an instructional design is a "blueprint" or a set of instructional-design rules. Designing instruction is rule-based. Instructional-design rules rest both on subject matter (concept, principle) and on the task to be performed (Dijkstra, 1986). Gulmans and Van der Sijde (1985) mention three criteria that instructional-design rules should meet (the same hold for managerial rules): deduction from theory; applicability in practice; completeness.
The third criterion (completeness) is implicit if the two others have been met. Deduction and applicability can only be accomplished by combining theoretical knowledge (in general: cognitive approaches to teaching) and empirical results (such as correlations between teacher behavior and student learning outcomes). Or, as recommended by Ryan and Anderson (1984), the design of experimental studies should be based on both a conceptual framework and on empirical evidence obtained from a variety of studies (i.e., experimental as well as correlational studies).
Also independent of the approach in the teacher training, the final test of the effectiveness lies in improving the learning outcomes of the students.
The nature of the evaluation of implementation is approach-dependent. One approach studies the implementation of the particular curriculum, while the other studies the implementation of the teacher training.

The research reported in this book studies the effects of teacher training - including subject-independent rules for instruction and managerial rules to extending the repertoire of teaching behaviors - on student and teacher learning outcomes. This chapter provides the theoretical and empirical basis for such a training.

CHAPTER 6

TEACHER TRAINING AND OBSERVATION INSTRUMENT

As described in the previous chapter, the content of the training is based on three main sources: first, the correlational research described in the Chapters 2 to 4, second, related research also conducted with the observe-correlate-experiment loop approach as the guiding principle, and third, general research on teacher training, classroom mangement, and instruction. The basic framework is described in the previous chapter. For ecological reasons the training is develloped particularly, for the users of the textbook Sigma. In the next section, we outline the content of the training, and describe the training procedure, and the observation instrument in relation to the training content.

1 THE CONTENT OF THE TRAINING

1.1 The Teaching Script

The leading idea behind the teacher training is the teaching script. This script consists of five different scenes: homework check, presentation, monitored practice, guided practice, and homework/tutoring.
Many combinations are possible, but in the training, after extended discussions with teachers and teacher trainers, we proposed one specific formulation of a teaching script. Every lesson should start with a review of homework, giving the teacher the opportunity to check for problems concerning the subject matter. This is necessary to conduct the presentation. When the teacher knows the problems, he can take them into account. After the presentation, he must check for clarity by giving the students an individual assignment. After the individual assignment (monitored practice), the teacher discusses a more extended and

difficult problem with the students (guided practice). The last phase of a lesson is giving the students homework and letting them start it in class, so he can tutor students with minor problems in the meantime. For each lesson phase or scene, we gave an indication of engagement time (see Table 6.1).

Table 6.1 An indication of time to spent in each of the lesson phases

Review of homework	5 to 10 minutes
Presentation of subject matter	10 to 15 minutes
Monitored practice	5 minutes
Guided practice	10 minutes
Homework assignments / tutoring	10 minutes
Total of time	45 minutes

The rationale behind this lesson schema is the correlation between time-on-task and student learning outcomes: r = .65 for the users of the textbook Sigma. A teacher maximizes time-on-task executing this particular script.

1.2 Classroom Management

The correlational study for the users of the textbook Sigma showed several interesting correlations between management behavior of the teacher and learning outcomes (see Table 6.2).

Table 6.2 Correlation coefficients between affective and cognitive learning outcomes and management behavior

Variable name	affective outcome	cognitive outcome
Control and discipline	-.09	-.54
Non-academic interactions	-.25	-.56
Teacher busy with management	-.26	-.69
Discipline remark to class	-.19	-.59
Discipline remark to student	.06	-.53

Table 6.2 reveals that the main trend is a negative correlation between management behavior of the teacher and learning outcomes. All teacher activities are curative, not preventive. In the classroom management model of the previous chapter, we emphasized the preventive nature of classroom management. This prevention was recommended to the teachers based on our cognitive interpretation of Kounin's theory.

Teachers should perform two parallel tasks, instruction and management. The teacher constantly watches the class for activities that might affect the ongoing lesson (instruction). We label this parallel activity *overlapping*. When the teacher "identifies" elements that might disturb instruction, he must act. Through his activity he shows he is *withit*. To show withitness a teacher can use the following techniques:
- look at a student or mention his/her name;
- rebuke a student (the right student at the right moment);
- propose an alternative behavior (e.g., call the student to the blackboard);
- let the student describe the desired behavior (ask what he is doing and what he is supposed to do).

A teacher should *focus on the group* and not on individual students. He must hold the attention of the group. If he cannot, he might for example, call on students in *turns*, or *give pay- attention signals* to the group ("This is important", "This will be part of the test next week").

Further, the students should be aware of their responsibility for what goes on in a lesson. The teacher should stimulate their learning process through *student accountability*. He can:
- ask goal-directed questions ("Jim, how far are you"?);
- check work ("Who has not finished assignment 5"?);
- have students answer other students' questions.

A lesson requires a *continuous signal*, a continuous stream of class activities, no abrupt changes and transitions, and a more or less constant pace.

1.3 Instruction

The previous chapter outlined the convergence in research from a quantitative and qualitative point of view. From this outline, we can formulate several instructional recommendations for teachers related to specific scenes in the teaching script.

Presentation The presentation of subject matter should follow a short review of relevant concepts familiar to the students (refresh prior knowledge), or a summary of relevant and familiar subject matter. When introducing new subject matter, the teacher should use explicit and concrete examples from daily life. Further, the teacher should use the textbook efficiently, e.g., by indicating the relevant and irrelevant parts. At the end of the presentation, the teacher summarizes the main points.

Monitored practice The presentation preceeds a short, clear, individual assignment that enables the teacher to check the effectiveness of his presentation. This assignment has to be checked, and the teacher should react positively towards the students. During guided practice, the teacher walks through the class and regularly checks students. After the students complete the assigment, the teacher must check their work (e.g., by asking for the solution and answers). The teacher must be positive and task-oriented.

Guided practice After the individual assignment, the teacher discusses a more complex problem with the class. The solution should be the result of a group process, not an individual student's. The teacher probes, redirects questions, etc.

Homework assignments/tutoring The final phase of a lesson is the homework assignment. Students should start on these in class and finish at home. While the students begin their homework assignments, the teacher has time to help individuals or small groups with particular problems identified in earlier phases of the lesson.

Homework check Although last in this outline, the homework check is the first part of a lesson. A teacher checks homework in one way or another. He must limit the amount of homework to what he can check the next lesson.

2 THE TRAINING PROCEDURE

The training lasts one day. A manual describes the content of the training (Van der Sijde, 1986b). We discussed the content with the participating teachers. The training session program was:

- discussion of the teaching script	1.5 hours;
- discussion of classroom management	1.5 hours;
- discussion of instruction	3.0 hours.

(The hours mentioned are estimates).

The trainer (one of the authors - VdS) introduced each discussion and put forward recommendations. Then the trainer asked the teachers to relate their own experiences with the particular subject and to consider feasibile. The trainer helped them agree on the usefulness of the recommendations in their own lessons, and in this way tried to influence their implementation.

3 THE OBSERVATION INSTRUMENT

The observation instrument has two parts: the general observation system TOOL, and the particular observation instrument TOOL-W.

Van der Sijde and Dirksen developed the TOOL-system, a methodology to design observation instruments on a small portable computer. The TOOL-system is a generic system used to develop different kinds of instruments. For details on this generic system, we refer to Van der Sijde and Dirksen (e.g. 1987a,b; 1988a,b).

The TOOL-W instrument measures the teacher activities that the training might influence. The first category is the lesson scenes. As described above, a lesson consists of a number of scenes (in TOOL-W, lesson phases) and these must be recorded:

- phase 1 - review of homework;
- phase 2 - presentation;
- phase 3 - monitored practice;
- phase 4 - guided practice;
- phase 5 - homework assignments and tutoring;
- phase 6 - all other lesson activities.

The second observation category is classroom management. This category includes the following observation codes:

- place- the teacher changes place in the class;
- interruption- the teacher interrupts the signal;
- pay-attention- the teacher gives pay-attention sign;
- turns- the teacher calls on students in turns;
- look- the teacher looks directly at a student;
- rebukes- the teacher rebukes a student;
- describes behavior- the teacher describes desired behavior;
- progress- the teacher asks for a progress report.

The third category, instruction, has the following codes:

- review- the teacher conducts a review;
- examples- the teacher uses an example;
- reference- the teacher makes a reference to the book;
- importance- the teacher indicates what is important;
- summary- the teacher summarizes;
- recall question- the teacher poses a recall question;
- high-level question- the teacher poses a high-level question;
- probe- the teacher probes;
- redirection- the teacher redirects a question;
- student question- the student asks teacher a question;
- good- the teacher gives a student feedback "good";
- good and rephrase- the teacher gives feedback "good" and rephrases the answer;

- wrong- the teacher gives feedback "wrong";
- wrong and good answer- the teacher gives feedback "wrong" and anwers the question himself.

The final category is direction of gaze. This category indicates alertness of the observers, and serves as an extra management category.

- class- the teacher's gaze is towards class;
- blackboard- the teacher's gaze is towards blackboard;
- paper- the teacher's gaze is towards paper or book;
- else- the teacher's gaze is elsewhere.

In TOOL-W, the observations are recorded every 5 seconds.

4 CORRESPONDENCE BETWEEN THE TRAINING AND TOOL-W

The correspondence between the recommendations in the training and the observation instrument is clearly illustrated in Figure 6.1.

TOOL-W	Training
Direction of gaze	
class	
blackboard	
paper	withitness
elsewhere	
Management place	
interruption	continuous signal
pay-attention	group alerting
turns	
look	
rebukes	
describes behavior	
progress	student accountability
Instruction	
review	review relevant subject matter
examples	use examples
reference	indicate importance
summary	summarize
recall question	pose recall questions
high-level question	pose higher-level questions
probe	use probes
redirection	redirects questions
student question	
good	give positive feedback
good and rephrase	
wrong	
wrong and good answer	
Lesson phases	
phase 1	lesson phase model
phase 2	
phase 3	
phase 4	
phase 5	
phase 6	

CHAPTER 7

IMPLEMENTATION OF A RESEARCH-BASED TEACHER TRAINING

This chapter attempts to trace how changes in the teaching script resulting from training affect the actual teaching behavior. As Duffy and Roehler (1985) state, teachers particularly resist complex, conceptual longitudinal changes; they more readily accept temporary or management routine changes. Gliessman (1985) disputes the observations of Duffy and Roehler (1985). He found that teachers could change their complex teaching behavior through conceptual methods of training. A conceptual method of training, by full understanding a teaching skill, can help to foster it. The two studies show that, the nature of the change aside, one can change teaching behavior.

Fullan & Pomfret (1977) give four reasons to study the training implementation. First, we do not know what has changed unless we can conceive and measure directly. The training manual contains recommendations for "effective" instruction and management. These recommendations have to be operationalized in observable behavior so that we can compare it to the pre-training behavior. We can use systematic observation (Leinhardt, 1975) tò measure the degree of implementation. If no observable change has occured, then the training could need revision or the recommendations may need better and more appropriate operationalizations.

Second, by investigating implementation directly, we can identify some of the most problematic aspects of bringing about change. By investigating implementation directly, e.g., through systematic observation, we discover which recommendations teachers can apply or implement, and which they cannot.

Third, neglecting to study implementation may result in its being ignored, or possibly, confused with other aspects of the change process. When we study an implementation process, we discover whether or not a teacher tries to use the

innovation (in this study, the training), succeeds in applying the training to his actual teaching behavior, or, for whatever reason, ignores the innovation.
Finally, unless we study implementation separately, it may be difficult to interpret learning outcomes and relate these to possible determinants. If the teacher does not implement the training in his actual teaching behavior, the training will have to be thoroughly revised. A second, experimental study must examine the influence of the teacher training on students' achievements and attitudes.
We must study implementation to determine if, in fact, any change has taken place, and to understand why change occurs or fails to occur (Fullan & Pomfret, 1977). Through systematic observation we study the effects of the teacher training on the actual teaching behavior.

1 METHOD

1.1 Subjects

Ten eighth grade mathematics teachers using the textbook Sigma voluntarily participated in this experiment. During the experiment, one teacher became ill for a longer period; we excluded the data on this teacher from the analysis. All the teachers were male, the mean number of years teaching experience was 11.3 (standard deviation: 6.2), and they worked at schools near the University of Twente.

1.2 Textbook

For this study we chose to use the textbook Sigma (Van Dop et al., 1979; Van Bemmelen et al., 1984; Cohen et al., 1979) (see also note 3.1). We chose it for ecological validity.

1.3 Observation Instrument TOOL-W

TOOL-W is an observation instrument implemented within the TOOL- system on a Canon X07 handheld microcomputer (Van der Sijde & Dirksen, 1987a,b; 1988a,b). The acronym TOOL stands for *"Toegepaste Onderwijskunde Observatie-instumentarium voor de Leeromgeving"* (Department of Education Observation Instruments for the Learning Environment). The TOOL-system is a flexible observation system installed on a small microcomputer (20 cm x 13 cm x 3 cm), and able to implement all kinds of observation instruments. TOOL-W is such an instrument, and it is described in detail in Chapter 6.

During an observation, every 5 seconds for 5 minutes, the observer enters one category from direction of gaze, one from lesson phase (optional), two from management (optional), and two from instruction (optional). After 5 minutes of observation, the observer has a 5 minutes rest period followed by another 5 minutes observation period. In general, a lesson consists of 5 observation periods of 5 minutes.

1.4 Training of Observers and Teacher Training

The training started with an explanation and discussion of the TOOL-W observation categories. Further, the training consisted of practical exercises using videotaped and audiotaped situations. During the final part of the training, we assessed interobserver agreement on most of the exercises by pairwise comparison of the observation codes. No formal estimation of the agreement was calculated). At the end of the training, this interobserver agreement was satisfactory.
The content of the teacher training and the training session procedure are discussed in Chapter 6.

1.5 Procedure

In the period March-April 1986 six trained observers sampled the observation data of six teachers during 6 lessons (two lessons per week) and collected the data of the three other teachers during three lessons (who joined the experiment two weeks later). The teachers visited our institute during their Easter-holiday for the one day in-service training. In the period following the Easter-holidays (April-May 1986) we again observed them for six lessons (one lesson per week). To inform the teachers about their teaching behavior in the period April-May (after the training), we sent them a letter within 24 hours with feedback on their lesson regarding the lesson phase model, classroom management, and instruction. After the experiment had been completed, the teachers again visited the institute for an informal gathering to evaluate the experiment. Seventeen lessons were observed by two observers to calculate the interobserver agreement.

1.6 Design and Data

This study employs a one group, pretest-posttest design. To study the implementation of the lesson phase model, we have constructed three variables. First, the L-score which is calculated according to the formula: L = 5 - the number of observed lesson phases; in the ideal lesson, L = 0. The second variable is F; F =

0 when the lesson starts with phase 1. Otherwise F = 1. For every lesson, a calculated S-score expresses the number of violations against the lesson phase model. An ideal lesson, according to the lesson phase model, starts with phase 1, followed by 2, 3, 4 and 5. Violations of the model are e.g. phase 4 following phase 1 (forward violation) or phase 1 following phase 3 (backward violation). The ideal lesson has an S-score of S = 0. Some violations are minor (from phase 4 to phase 3, or from phase 3 to phase 5), while others as major (from phase 5 to phase 2). S is the third variable to establish the degree of implementation of the lesson phase model.

Further, in TOOL-W a number of Kouninian variables are operationalized. We only report on the Kouninian variables in the results. *Withitness* is operationalized in low-inference variables. These low-inference variables are: the teacher watches the student (direction of gaze is toward the students), the teacher changes place (he takes another position in the classroom), the teacher looks at a particular student to indicate he has seen the pupil show inappropriate behavior, the teacher rebukes a student, and the teacher suggests alternative behavior to a student who shows inappropriate behavior. The observer registers the number of times the teacher performs such behavior before and after the training. *Withitness* is the standarized sum of the five low-inference behaviors mentioned. *Group alerting* is operationalized by the number of times the teacher alerts the class, e.g., with "this is important" or "pay attention". *Learner accountability* is operationalized by the number of times a teacher asks a student for a progress report (e.g., "who has completed ....", "who is not ready by now?").

The increase (or decrease) on management and instruction behavior is expressed as the effectsize. Effectsize E is defined as (X_A - X_B/sd_B), where X_A is the mean score after the training, X_B is the mean score before the training, and sd_B is the standard deviation of X_B (Glass, 1978).

2 RESULTS

2.1 Interobserver Agreement

We did not access interobserver agreement during the training. However, we made a hard copy of the data for every observation during the training period. Pairs of two or three observers compared and discussed differences, with each other and later with the experimenter. At the end of the training, the observers reached reasonable agreement, according to the experimenter. During the observation periods of the experiment, the mean interobserver agreement was calculated using Cohen's kappa (Cohen, 1962). The kappa (k) for the direction of gaze is: k = .50; for the lesson phases: k = .92; for the classroom management behavior: k = .73; and for the instruction behavior: k = .52.

2.2 Implementation of the Lesson Phase Model

In the training we discussed a lesson phase model and asked the teachers to implement this model. Based on the observed lessons we calculated the implementation scores of the lesson phase model and compared the scores before and after the training.

According to the lesson phase model, each lesson starts with phase 1. Before the training, 76% of the lessons started with phase 1 (F = 0.76), while after the training, 79% (F = 0.79) did. The difference is not significant: X_{22} = 0.23 (p >.10). Furthermore, each of the 5 phases has to be present in the lessons (Table 7.1).

Table 7.1 The percentage of lessons by number of lesson phases present in pre- and post-training observations.

	number of lesson phases present 5	4	3	2	1
Pre-training observations	1.6	17.7	35.5	38.7	6.5
Post-training observations	19.9	30.2	30.2	15.9	4.8

The mean L-score before the training is 2.3, and after the training, 1.6. After the training we see a significant change towards lessons with more lesson phases: X_{82} = 34.88 (p < .01). Every lesson includes violations of the model; before the training, the mean number of violations S is 4.92, standard deviation of 2.65, and after the training S is 4.06, with a standard deviation of 2.49. The difference is not significant t16 = -.46, p >.10.

During the training, we presented a model for the available time for each lesson phase. Table 7.2 presents the average amount of time the teachers spent on each lesson phase before and after the training.

Time spent on instruction increases from approximately 38.2 to approximately 43 minutes. We have extrapolated the figures in Table 7.2 for a complete lesson of 50 minutes; as mentioned before we systematically observed only 25 minutes of each lesson. Further, we see a significant decrease in time spent on "review of homework", and a significant increase in time spent on "presentation", "monitored practice" and "guided practice".

Table 7.2 Time (in minutes) engaged in each of the lesson phases, recommended time, and observed time before, and after the training

lesson phases	recommended	before training	after training	p*
Review of homework	5-10	22.4	16.6	< .10
Presentation	10-15	5.4	8.4	< .05
Monitored practice	5	3.4	7.2	< .05
Guided practice	10	2.2	4.4	< .05
Homework/tutoring	10	4.8	6.4	

*t-test, one tailed.

2.3 The Implementation of Classroom Management Behavior

The mean withitness-score and its standard deviation before the training are 0.02 and 2.41 respectively, and after the training 0.16 and 2.4 respectively; the effectsize of the training on withitness, E, is 0.06. The mean group alerting-score and its standard deviation before the training are -0.02 and 1.4 respectively, and afterwards -0.06 and 1.2 respectively; the effectsize of group alerting, E, is -0.01. Before the training, the mean score for learner accountability and its standard deviation are -0.001 and 1.0 respectively, and afterwards -0.02 and 1.0. The effect, E, is -0.02. The individual teacher withitness increased for 5 of the 9 teachers, and the same ratio occurred in group alerting- and learner accountability. Although classroom management in the training focussed on preventing classroom disorders, it also attempts to influence behavior that still interferes with instruction (e.g., rebuking) as well as behavior that does not (e.g., changing to another place in the classroom). The noninterfering teaching behavior has the largest effect: E = 0.67. Before the training the mean noninterfering behavior was 4.98, and after E = 6.69; t_{16} = 1.46, p ‹ .10. The effect of the interfering behavior: E = 0.42. Before the training, interfering behavior was 1.64, and after 1.38; t_{16} = 0.79, p ›.10, an insignificant difference.

2.4 Implementation of the Instruction Behavior

Instruction has three components: explanation, questions, and feedback. The recommendations regarding explanation concern the cues mentioned in Table 5.1. According to our recommendations, every lesson, as far as it concerns the phase "presentation", should start with a review of relevant concepts. Before the training, four teachers reviewed each lesson; after the training seven teachers. Students can also be asked to review. Before the training, the mean time per lesson that the teacher was engaged in review questions was 0.48 sec., and after the training 0.56 sec.; the effectsize E = .28. Explaining may also empoy the textbook; the mean amount of time per lesson that the teacher referred to the textbook was 0.14 sec. both before and after the training. Instead of referring to the textbook, the teacher can also indicate (e.g., on the blackboard) the important issues; the training did not change teaching behavior in this aspect. Before the training, the teachers spent far less time on the use of examples (0.05 sec. per lesson) than after (0.17 sec/lesson). The effect E = 1.75, and the difference is significant (t_{16} = 2.00, p .05). Presentation should end with a summary; two teachers summarized every lesson before the training, and after four teachers did.

Questions form a second component of explanation. After the training, the teachers used more high-level questions (E = 1.13; t_{16} = 2.10, p ‹ .05) and fewer recall or recognition questions (E = -0.18; t_{16} = -0.44, p ›.10). Further, they redirected more questions (E = 0.56; t_{16} = 0.16, p ›.10), and made fewer probes (E = -0.07; t_{16} = -0.17, p ›.10). Moreover, they decreased multiple probes (E = 0.42; t_{16} = -1.11, p ›.10), and increased high-level questions during "presentation" (E = 0.48; t_{16} = 0.89, p ›.10). The most appropriate place for high-level questions is the discussion, and during this phase we observed an increase of such questions (E = 1.76; t_{16} = 1.95, p ‹ .05).

The third component of explanation is feedback. For positive feedback, E = 0.17, and for negative feedback, E = 0.31.

3 DISCUSSION

The data on the use of time spent in a lesson phase show the training has changed behavior. Though all changes are in the desired (advised) direction, one of the lesson phases takes more time than advised (review of homework, although we observed a drastic decrease of about 6 minutes). Monitored practice took about 3 minutes before the training and about 7.2 minutes after the training (the advice was 5 minutes). Maybe teachers need more than 5 minutes to carry out all the tasks needed in monitored practice, especially checking the assignments. We observed an overall increase of effective time-on-task of approximately 5 minutes, from 38 minutes to 43 minutes. One lesson lasts 50 minutes, including changing classroom for the next lesson. On the whole, a substantial increase. Important in this context is that the teachers adopt the lesson phase model. Although we developed this model in discussion with teacher trainers, we did not know whether it would or could be implemented. The participating teachers could implemented the model, not perfectly on all occasions, but generally in one lesson at least. Their teaching script did actually change.

We conclude from Table 7.1 that the mode of the number of lesson phases before the training is 2, and from Table 7.2, that the lesson phases "review of homework" and "presentation" together comprise 73% of the time spent on instructional activities. After the training, the mode number of lesson phases is 3 and 4. An average lesson contains significantly more lesson phases, because of changes in the individual teachers' teaching scripts.
Although the teachers participating in our experiment were very experienced, they apparently were able and willing to implement the proposed behavior on of classroom management, especially on preventing classroom disturbances (withitness, learner accountability, and group alerting). They were not able to implement restoration of classroom disturbances, because the students elicit this kind of behavior. We observed the same classroom with the same students before and after the training and, the students were not trained at all. Nevertheless, we observed an effect on the classroom management behavior of the teachers.
The use of examples has a substantial effectsize. Walberg (1984), in his meta-analysis of research, mentions that "reinforcement" has the greatest effectsize, 1.17. Teachers did not review more after the training than before, but during "presentation", they asked more questions, both recall or recognition (questions on prerequisites) and high-level questions.
On the whole, the teachers posed fewer recall or recognition questions and more high-level questions than before the training. The most appropriate place for posing recall or recognition questions is during presentation, these questions are hard to formulate. High-level questions are easier to formulate, and we observed an increase of these questions (probably intended as recall or recognition questions).
Finally, we observed an increase of positive feedback (feedback "good" and feedback "good, and the teacher repeats answer") and a decrease of negative feedback (feedback "wrong" and feedback "wrong and teacher answers the question himself").
As a general trend in in-service training, Fullan (1986) mentions that "one-shot workshops are widespread and ineffective". The training (plus the manual) was a "one-shot workshop" and, in view of the results, effective. The training in this study proposed changes in the teaching script and in the activities in each of its scenes. For each recommendation based on this script, the training gave an explanation and evidence from research on teaching. In other words, the training was based on conceptual understanding of the content and its intention. Of course, we implemented the training to emphasize conceptual understanding. We had, however, another important reason related to the nature of the recommendations. All recommendations are practical, and most of them derive from

actual teaching behavior. Teachers combine information they receive (from researchers and educators) with what they already know, they restructure it, and make it fit their perception of teaching reality (Duffy and Roehler, 1985). During the training session, we saw that the teachers already practiced some of the recommendations, not systematically, but randomly. Thus, the recommendations already fit in with their knowledge structure, and they did not need to restructure the information.
Empirical evidence supports our conclusion that the training developed on the basis of the teaching script did change actual teaching behavior. Interestingly, the teachers participating in this study were no novice teachers (on the contrary!). This implies that when teachers voluntatily adopt an innovation, they are more likely to implement it, even if they are experienced teachers with routine teaching behavior. The results of this study do not call for changes in the teacher training manual, although two aspects of the training content need more explicite treatment during the one-day teacher training: the difference between recall or recognize and high-level questions, and the difference between guided and monitored practice. Despite these facts, the results indicate that the next study investigating the effect of teacher training on student achievement and attitude can ascribe an effect on these to this training.

CHAPTER 8

THE EFFECT OF A TEACHER TRAINING COURSE ON STUDENT LEARNING OUTCOMES

Teaching, the central process in education, attempts to transfer knowledge, attitudes, and skills to students. A teacher can use different kinds of models and strategies in teaching (see, e.g., Joyce & Weil, 1972). Studying the approaches mentioned by Joyce and Weil, we see that most models and strategies are based on educational views and theories that have not been investigated for classroom use. These theories assume that students in a classroom react the same as animal and human subjects in a laboratory, and researchers apply these learning theories to the classroom without further research. What actually happens in a classroom, and how learning and teaching take place in such a context, is still obscure.

Research according to the descriptive-correlational-experimental loop approach (Rosenshine & Furst, 1973) takes what actually happens in a classroom as its starting point. Using this approach, researchers can develop a theory based on experimental classroom research, without adhering to a specific theory to justify their design.

Fenstermacher (1976) demotes teaching a complex cognitive skill to a set of discrete teaching behaviors. This point of view criticizes training discrete units of teaching behaviors, such as microteaching. But it incorrectly assesses training developed as part of the descriptive-correlational-experimental loop approach. Such training is based on what teachers do in the classroom and, as a result, consists of a set of practical rules. Training such as that developed by Good and Grouws (1979) and Van der Sijde (1987), is not a set of independent rules reflecting discrete relationships between teaching behavior and student learning outcomes, but forms an instructional system that alters the concept of teaching, not just the behavior. In this way, teaching remains a complex skill. As a

number of training studies have already established, such training can improve student learning outcomes (Gage, 1985a,b).
The training, that is used in this study, is based on the teaching script (described in Chapter 5) with five scenes: presentation, guided practice, monitored practice, review of homework, and tutoring. For each of these five scenes, instructional and managerial rules have been formulated in order to execute the teaching script. Each rule expresses a cause-effect relationship, but, because the training is considered a conceptual whole, we cannot establish the effect of just one rule. Training does not reveal particular cause-effect relationships as Gage and Giaconia (1981) have proposed.
In order to interpret improvements in student learning outcomes resulting from teacher training, we conducted an implementation study (reported in Chapter 6). The study showed that teachers not only adopted, but actually implemented the teaching script in their teaching. The study concludes that training given to teachers can change their teaching behavior.
The study in this chapter investigates the improvement of student learning outcomes (achievement and attitude) as an effect of the teacher training.
The training studies reported by Gage (1985 a,b) used a nonequivalent control group design (Campbell & Stanley, 1966). Consequently, the teacher training accounts for differences in the improvement of student learning outcomes. Twelve out of thirteen studies support these findings. This study uses a slightly different design, not to investigate if a teacher training has an impact on student learning outcomes, but when it has the largest impact, i.e., before or after an observation period.

1 METHOD

1.1 Subjects

We asked about sixty teachers to participate in this study, and thirty-three teachers volunteered to do so. All teachers taught eighth grade mathematics using the textbook Sigma. We assigned the subjects to four conditions: condition one, thirteen teachers; condition two, eight teachers; condition three, six teachers; and condition four, six teachers, one of whom dropped out. The teachers in conditions one and two participated from September till March, while the teachers in conditions three and four only participated from December till March. The teachers came from in the vicinities of the State University of Groningen, the Technical University of Delft, and the Technical University of Eindhoven (all universities in the Netherlands).

In chapter 6 we described both the teacher training and the TOOL-W observation instrument.

1.2 Observers

Nine graduate students (in Education and Science) trained for about 40 hours to master the TOOL-W observation instrument. The three teams of observers were stationed at the Technical University of Delft (n = 3), the State University of Groningen (n = 4), and the University of Twente (n = 2) for the University of Eindhoven vicinity.

1.3 Training of Observers

Training started with an explanation and discussion of the TOOL-W observation categories. Further, it consisted of practical exercises using videotaped and audiotaped situations. The final part of the training involved three observations in the classroom. During the training, we assessed the interobserver agreement for most of the exercises by pairwise comparison of the observers' data.

1.4 Attitude Test

The attitude test was composed of items from the student questionnaire developed for the Second Mathematics Study of the International Association for the Evaluation of Educational Achievement (IEA). The original questionnaire (published in Pelgrum, Eggen, and Plomp, 1983) contained 84 items on six subjects: (1) mathematics at school; (2) mathematics as process; (3) mathematics and I; (4) mathematics and society; (5) gender and mathematics; (6) computers and calculators. We reanalyzed the data, using only Havo and Vwo data, in order to construct an attitude test investigating only the first three subjects. Factor analyses on the data of 1500 subjects revealed that the items within the subjects "mathematics at school", "mathematics as process", and "mathematics and I" loaded on two factors, which we can label "enjoying mathematics" and "self-image". Through item analysis, using the program "Reliability" in SPSS, we constructed two scales: enjoying mathematics (number of items is 13, Cronbach's alpha is 0.85, and examples of items are :"I like mathematics", "I want more mathematics") and self-image (number of items is 5, Cronbach's alpha is 0.85, and examples of items include: "I'll never be good at mathematics", "I have more trouble with mathematics than others do"). Each item had to be rated on a five-point-scale, and a score for each of the two subscales was calculated.

1.5 Achievement Test

We administered achievement tests to the classrooms of the participating teachers on three occasions (September 1985, December 1985, and April 1986). Each achievement test contained 30 multiple choice items constructed by the Dutch Central Institute for Test Development (CITO). The September achievement test (for classrooms of the teachers in experimental conditions 1 and 2) contained items on seventh grade subject matter. The December test (for classrooms of all participating teachers) covered subject matter from the period September-December, while the April test covered subject matter from the period January- April. The September test was the same for all classrooms; the teachers were asked to check whether the items in the test covered subject matter from the seventh grade; if not, these items eliminated (check for opportunity to learn). The December and April tests were slightly different for each classroom because of differences in subject matter covered in the respective periods (differences in place and type of school). The teachers also checked these tests for opportunity to learn.

1.6 Procedure

There are four experimental conditions. The teachers in condition one participated in a one-day training course at the University. Shortly after the training, we administered a student achievement and student attitude test to their classrooms. Trained observers then observed eight to ten lessons during a 2-month period (September-December) using the TOOL-W instrument. After the observations, we administered the same attitude test and a second achievement test, and, after another period of 2 months (January- April) the same attitude test again with a third achievement test. Condition two resembles condition one except that the teachers did not train before the observation period, but untill after the observations of their lessons (in January). The teachers in conditions three and four participated the December- April period. In December, we administered a student achievement and student attitude test to their classrooms. Only the teachers in condition three received a copy of the training manual by mail. In April, we also administered a student achievement and student attitude test to the classrooms of the teachers in conditions three and four. Conditions one and two are experimental conditions, while conditions three and four are control conditions. Different pairs of two observers observed 15 lessons to assess the interobserver agreement.

1.7 Design and Data

Figure 8.1 shows the design used in this study.

Condition 1	0	X1	0		0
Condition 2	0	X2	0		0
Condition 3			0	X3	0
Condition 4			0		0

Figure 8.1 Design of the study (0-administering of tests, X1, X2, X_3 - treatments)

In order to study implementation, we use the same variables as in the implementation study. To compare the increased scores on the achievement test over a period of time, we convert the mean classroom-score to a standard score using the overall mean of the group and the standard deviation.

2 RESULTS

2.1 Interobserver Agreement

During the training, we did not assess interobserver agreement, but made a hard copy of the data for every observers observation during the training period. Pairs of two or three observers compared the data and discussed the differences with each other, and afterwards with the experimenter. At the end of the training, the observers had reached reasonable agreement. During the observation period of the experiment, we calculated the mean interobserver agreement. We assessed the interobserver agreement using Cohen's kappa (Cohen, 1960). The mean kappa (k) for the direction of gaze, was: $k = 0.48$; for the lesson phases: $k = 0.98$; for classroom management: $k = 0.84$; and for instruction: $k = 0.59$.

2.2 The Implementation of the Lesson Phase Model

The averages in Table 8.1 show that the teachers in the first condition spent significantly more time on monitored practice: $t_{19} = 1.60$ ($p < .10$), and on guided practice: $t_{19} = 1.30$ ($p < .10$) than the teachers in the second condition.

Table 8.1 Time (in minutes) engaged, in each of the lesson phases: recommended time, observed in condition 1, and observed in condition 2

lesson phases	recommended	Condition 1	Condition 2
Review of homework	5-10	16.50	19.40
Presentation	10-15	9.76	10.28
Monitored practice	.5	8.90	5.10*
Guided practice	10	3.50	1.76*
Homework/tutoring	10	5.20	6.33

*t-test, one tailed, p < .10.

During the training, we emphasized the presence of five lesson phases in each lesson. Table 8.2 shows the percentage of the number of lesson phases for the first two conditions. Comparing the data, the first condition yields a significant difference: $X^5_2 = 27.16$, $p < .001$. The conditions do not reveal a difference with respect to the F-variable (starting a lesson with a review of homework). The mean number of violations against the model (S- variable) and the standard deviation for teachers in condition one are 3.12 and 3.40 respectively and for teachers in condition two, 4.00 and 4.69 respectively. The difference is significant: $t_{189} = -1.49$, $p < .10$.

Table 8.2 The percentage of lessons by number of lesson phases present in pre- and post-training observations.

	number of lesson phases present				
	5	4	3	2	1
Condition 1	1.6	17.7	35.5	38.7	6.5
Condition 2	19.9	30.2	30.2	15.9	4.8

2.3 Implementation of Classroom Management Behavior

The mean withitness-score and its standard deviation of the teachers in condition one are 0.26 and 2.85 respectively, and of the teachers in condition two -0.42 and 2.00. Comparing these conditions for withitness, the difference is not significant; E = 0.34. The mean group alerting-score of teachers in condition one and its standard deviation are -0.05 and 1.02 respectively, and of teachers in condition two, 0.31 and 0.98. Comparing the conditions for group alerting, the difference is not significant; the effectsize, E = .037. The mean score for learner accountability of teachers in condition one and its standard deviation are 0.30 and 1.08 respectively, and of teachers in condition two, -.037 and 0.73. The difference between the conditions is significant: MWU = 26.5, $p < .10$; the effectsize, E = 0.92.

2.4 Implementation of Instruction Behavior

Instruction is divided into three components: explanation (presentation of subject matter), questions, and feedback. The means in Table 8.3 show that reviewing (at the start of presentation) has the largest effectsize (E = 1.00), while summarizing (at the end of presentation) has a negative effectsize (E = -0.43). The training has a small effect on the time engaged in recall and recognize questions (E = 0.15), and on positive feedback (E = 0.60).

Table 8.3 The mean number of 5 seconds intervals, average time and standard deviations engaged in particular instruction activities by the teachers in both conditions.

	Condition 1			Condition 2		
	mI*	aT**	sd	mI	aT	sd
EXPLANATION						
-review	.55	.5	.50	.25	.2	.30
-ask about prerequisites	2.75	2.3	1.95	2.55	2.1	1.60
-use textbook	.55	.5	.45	.30	.3	.40
-indicate importance	.65	.5	.30	.60	.5	.55
-use examples	.25	.2	.25	.20	.2	.25
-summarize	.30	.3	.35	.60	.5	.70
QUESTIONS						
-recall	10.30	8.9	4.30	9.65	8.0	4.20
-higher level	.55	.5	1.00	.50	.4	1.10
FEEDBACK						
-positive	1.04	1.0	.65	.76	.6	.60
-negative	.40	.3	.25	.49	.4	.39

mI*: mean number of 5 seconds intervals;
aT**: average time in minutes.

2.5 Effects of the Training on Student Achievement

Table 8.4 shows the mean scores on the students' achievement test for all conditions.

Table 8.4 Mean classroom scores and their standard deviations on achievement tests administered in September, December and April.

	Period of test administration		
	September	December	April
Condition 1	.622 (.073)	.524 (.159)	.466 (.084)
Condition 2	.609 (.080)	.595 (.190)	.519 (.164)
Condition 3	-	.493 (.174)	.420 (.107)
Condition 4	-	.609 (.154)	.532 (.137)
Mean	.616 (.074)	.549 (.167)	.481 (.121)

The scores for conditions one and two on the achievement test do not differ significantly either in September, December, or April. Further, after the analysis of variance with repeated measures on achievement, we see a main effect on repeated measures, $F(2,38) = 10.95$, $p < .01$. This shows that achievement decreases significantly during the year.
Comparing the increase in mean classroom score on achievement in the period September-December, we see no significant difference between the conditions. Nor do the two experimental conditions show a significant increase in scoring for the period December-April. An increase in mean classroom score on achievement is significant for the period September-April, in favor of the second condition: MWU = 27, $p < .10$. There are no significant differences between the conditions on the December tests (the control conditions are considered together on this test), nor are there on the April tests (the control conditions are now considered separately). On comparison of the increase in scoring in the December-April period, the conditions do not differ significantly, nor do the experimental and control conditions.

2.6 Effects of the Training on Student Attitude

2.6.1 Effects of the training on student attitude: enjoying mathematics

Table 8.5 shows the mean scores on student attitude test, subscale "enjoying

mathematics", for all conditions. On comparison of the scores on the subscale "enjoying mathematics", conditions 1 and 2 do not differ significantly, either in September, December or April.
On comparison of the increase in mean classroom score on the subscale "enjoying mathematics", we see no significant difference between the conditions in the period September- December, or the period December-April.
The conditions do not differ significantly on the December test (the control conditions are considered together), nor on the April test. On comparison of the increase in the December-April period, the experimental and control conditions do not differ significantly.

Table 8.5 Mean classroom scores and their standard deviations on the attitude test 'enjoying mathematics', administered in September, December, and April.

	Period of test administration		
	September	December	April
Condition 1	21.36 (1.79)	21.05 (2.00)	20.93 (1.76)
Condition 2	21.43 (1.32)	21.60 (1.13)	21.67 (1.19)
Condition 3	-	21.21 (1.75)	20.20 (1.75)
Condition 4	-	21.97 (1.31)	21.82 (1.38)
Mean	21.39 (1.59)	21.36 (1.64)	21.12 (1.61)

2.6.2 Effects of the training on student attitude: self-image

Table 8.6 shows the mean scores on student attitude test, subscale "self-image", for all conditions.
On comparison of the scores on the attitude test, subscale "self- image", conditions 1 and 2 do not differ significantly, in either September, December, or April.
On comparison of the increase in mean classroom score on the subscale "self-image" in the period September-December, we see no significant difference between the conditions, nor do they differ significantly in the periods December-April, or September-April.
The conditions do not differ significantly on either the December, or the April test. On comparison of the increase in scoring in the December-April period, the conditions do not differ significantly, nor do the experimental and control conditions. However, analysis of variance with two conditions (experimental and

control) and repeated measures on self-image reveal a significant interaction effect: $F(1,30) = 4.83, p < .05$.

Table 8.6 Mean classroom scores and their standard deviations on the attitude test 'self-image', administered in September, December, and April.

	Period of test administration		
	September	December	April
Condition 1	17.93 (1.66)*	18.05 (1.74)	18.85 (2.42)
Condition 2	18.04 (.90)	18.34 (.84)	19.04 (1.51)
Condition 3	--- ---	18.21 (1.74)	17.78 (2.10)
Condition 4	--- ---	18.75 (1.31)	18.77 (1.03)
Mean	17.97 (1.39)	18.26 (1.45)	18.69 (1.96)

* standard deviation

3 DISCUSSION

The results of the interobserver agreement show a reasonable agreement between the observers, meaning that the agreement is not coincidental. The observations are reliable and we can draw conclusions from the data gathered in this experiment. The results show that the classroom means of the teachers in conditions one and two do not differ significantly in either attitude or achievement scores. This holds for the September as well as for the December and April tests. Conditions one, two, three, and four show the same results on the December and April tests. From these results, we can deduce that the observer's presence did not affect the classroom results (no Hawthorne effect).
We studied the increase in scores from the September to the December and April tests to assess the effectiveness of the training. Comparing the difference in increase in mean classroom scores on the achievement test and the attitude tests between the teachers in the first and second experimental conditions, we find no significant difference in the period September-December. Remarkably, no other training study has reported such a result. As in all other training studies, the first part of the experiment is a nonequivalent control group design, and although the training has not effected an increase on the achievement and attitude scores, it did effect the teaching behavior. As we see from the results of the classroom observations, the mean effect size of training, calculated as the mean over all recommendations, is approximately 0.33.

Further, the teachers in the first condition implemented the lesson phase model evident from the lower value of the F- and L- variables. We see no significant difference for the F-variable, which means that most teachers, regardless of the experimental conditions, start their lessons with homework check. The teachers in the first condition also spend time on monitored practice, while the teachers in the second condition spend significantly less time on this phase. The same holds for guided practice. The participating teachers indicated that the discrimination between guided and monitored practice, although in theory rather clear, is in practice hard to make.
If we group these two forms of practice together, we see that the teachers in the first condition spend approximately 12.5 minutes per lesson on practice, while the teachers in the second condition spend approximately 7 minutes; the lesson phase model advises approximately 15 minutes. We conclude that these differences in teaching behavior reflect differences in the teaching script.
The implementation study shows that the training influences teaching behavior and that the mean effectsize was approximately 0.27. If this implementation study had not been performed, we could not establish a realistic measure of implementation on the basis of this study: we would overestimate. The effectsize in the implementation study estimates the degree of implementation, because the teachers in the implementation study acted as their own control group, while those in this training study did not. The difference between the teachers in condition one and condition two arises from different teaching scripts and not from the training. Because we performed an implementation study, we can reach a more sophisticated conclusion: the implementation study showed that the training could be implemented, so we need not investigate implementation in this study. The teachers in experimental condition one adhere to the proposed teaching script more than those in experimental condition two (they were unaware of the content of the training and the particular teaching script); the teachers in experimental condition one adopted and actually practiced the teaching script. Although the teachers in the first experimental condition adhered to the teaching script, the changes in their actual teaching behavior did not effect an increase in test scores in the period September-December. Probably a period of approximately 2.5 months is too short. The teachers in the first experimental condition had to discover themselves to what degree they already adhered to the new teaching script. The increase on the achievement test scores differs significantly in September-April and favors the second experimental condition. We found no effect on the increase on the attitude tests scores. This means that the most effective condition places training after the observation period. There is an explanation for this effect. The teachers in the first condition participated in the training. The teachers may have found many of the recommendations and

rules plausible and familiar, because they express what a good teacher should do. We allowed the teachers in the first condition to combine what they do with what they should do in the classroom. In other words, these teachers did not know to what extent their teaching script resembled the teaching script proposed in the training. Because of this possible confusion, the teacher received no indications for changes (no data of classroom observation were available). The teachers in the second condition, however, distinguished their actual teaching behavior from desired teaching behavior. All observation data were available and used during the training: the teachers were confronted with the results of the classroom observations. In other words, these teachers received information on how their teaching script resembled the proposed teaching script, and as a result, we could give them concrete indications for change. Probably the most effective design gives training on two occasions: first, before a period of observations, and second, the same training, but tailored, after two or three months. A number of training studies had a second training session, but not in the sense meant above. In the training studies by Anderson et al. (1979) and Good and Grouws (1979) the period of time between the first and second training sessions was only one week.

When we compare the teachers in the third and fourth conditions on the increase in test scores in the December-April period, we see no significant differences. This means that mailing the teachers the training manual did not affect teaching behavior. Coladarci and Gage (1984) obtained the same results. Their design resembles the design in this study except that the length of time between the tests in their study was about one school year, and in the present study, 2.5 months. Coladarci and Gage found no significant differences. A difference between the present study and that of Coladarci and Gage (1984) is, that they asked the teachers to participate in the study: upon agreement, the researchers mailed manuals to the teachers in this experimental group. In our study, the teachers in the third condition received the manual and were asked to read and practice it. We did not check on reading the manual. Coladarci and Gage (1984) called this a "minimum intervention". Even with less minimal intervention, e.g., using a design with three conditions - face- to-face training (the second condition), a training manual alone (the third condition), and neither manual nor meetings (the fourth condition), we see no significant difference.

The minimum intervention in this study has been clearly too minimal regarding the degree of participation and time. A period of 2.5 months (including holidays) may be too short to influence observable student learning outcomes; however, it is not too short to notice changes in teaching behavior. Influenced by the training a teacher changes his teaching script and behavior. These changes take time, although feedback on the actual teaching behavior can shorten this. Comparing

the experimental conditions with the control conditions, we see an interaction between condition and repeated measures of self-image. As a result of the training, the mean self-image of students of the teachers in the experimental conditions improves more than the mean self-image of the students of the teachers in the control conditions.

This study, once again, establishes that even a short teacher training course can successfully change the teachers' teaching script, and, subsequently, their teaching behavior, which in turn influences student achievement.

REFERENCES

Abelson, R. (1981). Psychological status of the script concept. *American Psychologist, 36,* 715-729.

Allen, D.W., Ryan, K. (1969). *Microteaching.* Reading: Addison-Wesly.

Anderson, L.M., Evertson, C.M. & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79,* 193-223.

Anderson, J.R., Greeno, J.G., Kline, P.J. & Neves, D.M. (1981). Acquisition of problem-solving skill. In Anderson, J.R. (Ed.), *Cognitive skills and their acquisition.* Hillsdale: Lawrence Erlbaum Associates.

Arlin, M. (1979). Teacher transition can disrupt timeflow in classrooms. *American Educational Research Journal, 16,* 42-56.

Atwood, R. (1983). *The interacting effects of task form and activity structure on students' task involvement and teacher evaluations.* Paper presented at the Annual Meeting of the AREA, Montreal.

Bartlett, F.C. (1932). *Remembering: a study in experimental and social psychology.* London: Cambridge University Press.

Berliner, D.C. (1986). In persuit of the expert pedagogue. *Educational Researcher, 15* (7), 5-13.

Berliner, D.C. & Carter, K.J. (1986). *Differences in processing classroom information by experts and novices.* Paper presented at the Conference of the International Study Association on Teacher Thinking, Leuven.

Bloom, B.S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.

Bloom, B.S. (1977). Time and learning. *American Psychologist, 29,* 682-688.

Boekaerts, M. (1978). *Towards a theory of learning based on individual differences.* Tilburg: Katholieke Hogeschool Tilburg.

Borg, W. (1975). *Instructor handbook for the protocol modules on classroom management.* Logan: Utah State University.

Brophy, J.E. & Evertson, C.M. (1974). *Process-product correlations in the Texas Effectiveness Study: Final Report.* Research report no. 74-4, Austin, Texas: Research and Development Center for Teacher Education, University of Texas.

Brophy, J.E. (1979). *Advances in teacher effectiveness research.* East Lansing: Institute for Research on Teaching.

Brophy, J.E. (1980). *Recent research on teaching.* East Lansing: Institute for Research on Teaching.

Brophy, J.E. & Good, T.L. (1987). Teacher behavior and student achievement. In M.C. Wittrock, (Ed.), *Handbook of research on teaching, third edition.* New York: MacMillan.

Burn, B.B. & Hurn, C.J. (1982). *An analytical comparison of educational systems*. National Commission on Excellence in Education. Washington, D.C.: U.S. Department of Education.
Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi- experimental designs for research*. Chicago: Rand McNally.
Clark, C.M. & Peterson, P.L. (1986). Teachers' thought processes. In M.C. Witrock (Ed.), *Handbook of research on teaching, third edition*. New York: MacMillan.
Cohen, K.H. & Van der Houven, F.D.A. (1979). *Sigma 2MH*. Groningen: Wolters-Noordhoff.
Cohen, E.G., Intilli, J.A.K. & Robbino, S.H. (1979). Task and authority: a sociological view of classroom management. In Duke, D.L. (Ed.), *Classroom management*. Chicago: University of Chicago Press.
Cohen, J. (1962). A coefficient of agreement for nominal scales. *Eductional and Psychological Measurement, 20*, 37-46.
Cohen, L.A. (1987). Diet and Cancer. *Scientific American, 257*, 42- 48.
Coladarci, T. & Gage, N.L. (1984). Effects on a minimal intervention on teacher behavior and student achievement. *American Educational Research Journal, 21*, 539-555.
Crawford, J. & Gage, N.L. (1977). Developing an research-based teacher education program. *California Journal of Teacher Education, 4*, 105-123.
Crawford, J., Gage, N.L. & Stayrook, N. (1987). *An experiment on teaching in the third grade. (Research Memorandum)*, Program on teacher effectiveness, Center for educational research at Stanford. Stanford: SRI.
Creton, H. & Wubbels, T. (1984). *Ordeproblemen bij beginnende leraren: een analyse van de ordeproblemen bij beginnende leraren in het voortgezet onderwijs*. Utrecht: Rijksuniversiteit Utrecht.
Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psichometrika, 16*, 297-334.
Cuban, L. (1984). *How Teachers Taught: Constancy and Change in American Classrooms 1890-1980*. New York: Longman.
Dijkstra, S. (1986). De psychologie van het leren en de vormgeving van instructie. In Dijkstra, S. & Span, P. (Eds.), *Leerprocessen en instructie*. Lisse: Swets & Zeitlinger.
Doyle, W. (1986). Classroom organization and management. In M.C. Wittrock (Ed.), *Handbook of research on teaching, third edition*. New York: MacMillan.
Duffy, G.G. & Roehler, L.R. (1985). *Constraints on teacher change*. East Lansing: Michigan State University. Institute for Research in Teaching.
Duffy, G.G., Roehler, L.R., Meloth, M.S. & Vavrus, L.G. (1986). Conceptualizing instructional explanation. *Teaching and Teacher Education, 2*, 1-15.

Duke, D.L. (1979). Editorial preface. In Duke, D.L. (Ed.), *Classroom management*. Chicago: University of Chicago Press.
Dunkin, M.J. & Biddle, B.J. (1974). *The Study of Teaching*. New York: Holt, Rinehart & Winston.
Emmer, E.T. & Evertson, C.M. (1981). Synthesis of research on classroom management. *Educational Leadership, 38*, 342-347.
Epstein, A.S., Shulman, L.S. & Sprafka, S.A. (1978). *Medical Problem Solving*. Cambridge: Harvard University Press.
Evertson, C.M., Emmer, E.T. & Brophy, J.E. (1980). Predictors of effective teaching in junior high mathematics classrooms. *Journal for Research in Mathematics Education, 11*, 167- 178.
Evertson, C. & Emmer, E. (1982). Effective management at the beginning of the schoolyear in junior high classes. *Journal of Educational Psychology, 74*, 485-498.
Fenstermacher, G.A. (1976). Philosophical consideration of recent research on teacher effectiveness. *Review of Research in Education, 6*, 157-185.
Fisher, C.W., Filby, N.A., Marliave, R., Cahen, L.S., Dishaw, M.M., Moore, J.E. & Berliner, D.C. (1978). *Teaching behaviors, academic learning time, and student achievement: Final report of Phase III-B, Beginning Teacher Evaluation Study. (Technical Report V-1)*. San Francisco: Far West Laboratory for Educational Research and Development.
Fullan, M. (1986). School focussed inservice education. In D. Hopkins (ed.), *Inservice training and educational development*. London: Croom Helm.
Fullan, M. & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research, 47*, 335-397.
Gage, N.L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.
Gage, N.L. (1979). *The causal connection between teaching practices and student achievement: Recent two-stage experiments with teaching practices derived from correlational studies*. Stanford.
Gage, N.L. & Giaconia, R. (1981). Teaching practices and student achievement: causal connection. *New York University Education Quarterley, 12*, 2-9.
Gage, N.L. (1984). What do we know about teaching effectiveness? *Phi Delta Kappan*, 66, 87-93.
Gage, N.L. (1985a). *Hard Gains in the Soft Sciences: The case of pedagogy*. Bloomington: Phi Delta Kappa's Center on Evaluation, Development and Research.
Gage, N.L. (1985b). *What educators ought to know about research on teaching*. Paper prepared for the Doctor Residency Seminar, College of Education, University of South Florida, on October 12.

Gagne, R.M. (1976). The learning basis of teaching methods. In N.L. Gage (Ed.), *The psychology of teaching methods*. 21-43, Chicago: NSSE.

Galambos, J.A. & Rips, L.J. (1982). Memory for routines. *Journal of Verbal Learning and Verbal Behavior, 21*, 260- 281.

Gilbert, J.P. & Mosteller, F. (1972). The urgent need for experimentation. In Mosteller, F. & Moynihan, D.P. (Eds), *On equality of educational opportunity*. New York: Vintage Books.

Glaser, R. (1976). Components of a psychology of instruction: towards a science of design. *Review of Educational Research, 46*, 1-24.

Glass, G.V. (1978). Integrating findings: The meta-analysis of research. In L.S. Shulman (Ed.), *Review of research in education*, 351-379.

Gliessman, D.H. (1985). *A study of change in complex teaching skills*. Paper presented at the Annual Meeting of the AERA, Chicago.

Good, T.L., Biddle, B.J. & Brophy, J.E. (1975). *Teachers make a difference*. New York: Holt, Rinehart & Winston.

Good, T.L. & Grouws, D.A. (1977). Teaching effects: A process- product study in fourth-grade mathematics classrooms. *Journal of Teacher Education, 28*, 49-54.

Good, T.L. & Grouws, D.A. (1979). The Missouri mathematics effectiveness project: An experimental study in fourth grade classrooms. *Journal of Educational Psychology, 71*, 355-362.

Good, T.L., Grouws, D.A. & Ebmeier, H. (1983). *Active Mathematics Teaching*. New York: Longman.

Good, T.L., Slavings, R.L., Harel, K.H. & Emerson, H. (1987). Student Passivity: A study of question asking in K-12 classrooms. *Sociology of Education*, 181-199.

Goodlad, J. (1983). A study of schooling: Some findings and hypotheses. *Phi Delta Kappan*, 465-470.

Goodlad, J. (1984). *A place called school*. New York: McGraw-Hill.

Hartley, J. Trueman, M. (1982). The effect of summaries on the recall of information form prose: Five experimental studies. *Human Learning, 1*, 63-82.

Gulmans, J. & Van der Sijde, P.C. (1985). *Concepties over concepten*. Enschede, Toegepaste Onderwijskunde, IST-MEMO-85- 02.

Hol, D. (1983). *Routinematig gedrag van leerkrachten tijdens het onderwijzen*. Groningen: Vakgroep Onderwijskunde.

Irving, O. & Martin, J. (1982). Withitness: The confusing variable. *American Educational Research Journal, 19*, 313- 319.

Jamison, D., Suppes, O. & Wells, S. (1974). The effectiveness of alternative instructional media: A survey. *Review of Educational Research, 44*, 1-68.

Johnson, P.E., Duran, A.S., Haasselbrock, F., Moller, J. Prietula, M., Feltovich, P.J. & Swanson, D.B. (1982). Expertise and error. *Cognitive Science, 6*, 235-283.

Joyce, B. & Weil, M. (1972). *Models of teaching*. London: Prentice-Hall.

Kameenui, E.J., Carnine, D.W., Darch, C.B. & Stein, M. (1986). Two approaches to the development phase of mathematics instruction. *Elementary School Journal, 86*, 633-650.

Kieviet, F.K. (1972). *Microteaching als methode in de opleiding van leerkrachten*. Vaassen: Van Walraven.

Kieviet, F.K. (Ed.). (1978). *Nieuwe methoden in de opleiding van onderwijsgevenden*. Groningen: Wolters-Noordhoff.

Kounin, J. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.

Kounin, J. & Doyle, P.H. (1973). Degree of continuity of a lesson's signal system and the task involvement of children. *Journal of Educational Psychology, 67*, 159-164.

Kounin, J. & Gump, P. (1974). Signal systems of lesson settings and the task related behavior of preschool children. *Journal of Educational Psychology, 66*, 554-563.

Krammer, H.P.M. (1984). *Leerboek en leraar*. Harlingen: Flevodruk, Stichting voor Onderzoek van het Onderwijs.

Leinhardt, G. & Smith, D. (1985). Expertise in mathematics instruction: subject matter knowledge. *Journal of educational psychology, 77*, 247-271.

Leinhardt, G. (1986a). Expertise in mathematics teaching. *Educational Leadership, 43*, 28-33 (April).

Leinhardt, G. (1986b). Math lessons: a contrast of novice and expert competence. In J. Lowyck (Ed.), *Proceedings of the third congress of the International Study Association on Teacher Thinking*. Leuven: University of Leuven.

Leinhardt, G. & Greeno, J.G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*, 75-95.

Leith, G.O.M. (1983). Whatever happened to programmed instruction, mastery learning, and microteaching ...? Some reflections on a neglected area of educational technology. In *Aspects of Educational Technology, Volume XVI, Improving Efficiency in Education and Training*. New York: Kogan Page.

Levin, T. (1977). *Trends and findings in classroom environment research*. Tel-Aviv, Israel: School of Education, Tel-Aviv University.

Lifton, R.J. (1986). *The Nazi Doctors/Medical Killing and the Psychology of Genocide*. New York: Basic Books.

Lowyck, J. (1978). *Procesanalyse van het onderwijsgedrag*. Studie van de cognitieve variabelen in het voorbereidings- en uitvoeringsgedrag van ervaren leerkrachten: Literatuurstudie en descriptief-empirisch onderzoek. Proefschrift, Leuven: Katholieke Universiteit, Faculteit der Psychologie en Pedagogische Wetenschappen, Afdeling Didactiek en Psychodiagnostiek.

Lysakowski, R.S. & Walberg, H.J. (1981). Classroom reinforcement and learning: a quantitative synthesis. *Journal of Educational Research, 75*, 69-77.
Lysakowski, R.S. & Walberg, H.J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal, 19*, 559- 578.
McDermot, R.C. (1976). *Kids make sense: An etnographic account of the interactional management of succes and failure in one first-grade classroom*. Ph.D. Thesis, Stanford University.
McDonald, F.J. & Elias, P.J. (1976). *Executive Summary Report: Beginning Teacher Evaluation Study, phase II, 1973-1974*. Princeton: Educational Testing Service.
Merrill, M.D. & Boutwell, R.C. (1973). Instructional development: Methodology and research. In F.N. Kerlinger, (Ed.), *Review of research in education*. Itasca: F.E. Peacock Publishers.
Merrill, D.M. (1983). Component display theory. In: C.M. Reigeluth, *Instructional design theories*. Hillsdale: LEA.
Minsky, M. (1975). A framework for representing knowledge. In: P.H. Winston, *The psychology of computer vision*. New York: McGraw-Hill.
Neter, J. & Wasserman, W. (1974). *Applied linear statistical models*. Homewood: Irwin.
Pelgrum, W.J., Eggen, Th. & Plomp, Tj. (1982). Het tweede IEA wiskundeproject. In Van Dorp, C., en Pilot, A. (Eds.), *Aspecten van onderwijsresearch*. Lisse: Swets & Zeitlinger.
Pelgrum, W.J., Eggen, Th. & Plomp, Tj. (1983). Tweede wiskundeproject: *Analyse van uitkomsten, leerstofaanbod en resultaten*. Enschede: Toegepaste Onderwijskunde.
Plomp, Tj. (1982). *Onderwijskundige technologie: enige verkenningen*. Enschede: TH Twente, oratie.
Plomp, Tj., Wolde, J. van de, (1984). Een algemeen model voor onderwijskundig probleemoplossen. In E. Warries, (Red.). *Ontwerpmethodologie voor onderwijsproblemen*. Enschede: Technische Hogeschool Twente, Onderafdeling der Toegepaste Onderwijskunde.
Popper, K.R. (1972). *The poverty of historicism*. London: Routledge & Kegan Paul.
Redfield, D.L. & Rousseau, E.W. (1980). *A meta-analysis of experimental research on teacher questioning behavior*. Tucson, University of Arizona.
Reigeluth, C.M. (1983). Instruction: what is the discipline like. In Reigeluth, C.M. (Ed.), *Instructional design theories and models*. Hillsdale: Lawrence Erlbaum Associates.
Robitaille, D.F. (1975). Criteria for assessing the effectiveness of teachers of

secondary school mathematics. *Journal for Research in Mathematics Education. 6*, 77-87.
Rosenshine, B. & Furst, N. (1971). Research on teacher performance criteria. In B.O. Smith (Ed.), *Research on Teacher Education*. (37-72).
Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Ed.), *Second Handbook of Research on Teaching*. Chicago: Rand McNally.
Rosenshine, B.V., (1986). Synthesis of research on explicit training. *Educational Leadership, 43*, 60-69.
Rosenshine, B.V., & Stevens, R. (1986). Teaching Functions. In M.C. Witrock (Ed.,), *Handbook of Research on Teaching*, Third Edition, New York: Macmillan.
Rummelhart, D.E. (1977). Understanding and summarizing short stories. In Laberg, D. & Samuels, S.J. (Eds.), *Basic Processes in Reading*. Hillsdale: Lawrence Erlbaum Associates.
Ryan, D.W. & Anderson, L.W. (1984). The IEA Classroom Environment Study: Promises and pitfalls. *Evaluation in Education, 8*, 87-94.
Schank, R., (1982). *Reading and understanding: teaching from the perspective of artificial intelligence*. Hillsdale: Lawrence Erlbaum Associates.
Schank, R. & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale: Lawrence Erlbaum Associates.
Shavelson, R.J. & Stern, P. (1981). Research on teacher's pedagogical thoughts, judgements, decisions and behavior. *Review of Educational Research, 51*, 455-498.
Sirotnik, K.A. (1984). An inter-observer reliability study of a modified SRI Observation System. *Journal of Classroom Interaction, 19*, 28-38.
Snippe, J., & Harskamp, E. (1986). *Inservice training als determinant van curriculumimplementatie*. Paper gepresenteerd tijdens ORD'86 te Utrecht.
Soar, R. (1973). *Final Report, Follow through classroom process measurement and pupil growth. (1970-1971)*. Gainsville, Fla.: Institute for the Development of Human Resources, University of Florida.
Soar, R. & Soar, R.M. (1979). Emotional climate and management. In Peterson, P.L. & Walberg, H.J. (EDS.), *Research on teaching*. Berkeley: McCutchan.
Stallings, J.A. & Kaskowitz, D. (1974). *Follow through classroom observation evaluation, 1972-1973*. Menlo Park, Ca.: Stanford Research Institute.
Stallings, J.A. (1977). *Learning to look: A handbook on classroom observation and teaching models*. Belmont, Ca.: Wadsworth.
Stallings, J.A., Needels, M. & Stayrook, N. (1979). *How to change the process of teaching basic reading skills in secondary schools: Phase II and III, Final report*. Menlo Park, Ca.: SRI International.

Tennyson, R.D. & Cocchiarella, M.J. (1986). An empirically based instructional design theory for teaching concepts. *Review of Educational Research, 56*, 40-71.
Tomic, W., Weesie, C.A.M. (1982). Lesgedrag van wiskundeleraren. In C. van Dorp en A. Pilot (Red.), *Aspekten van Onderwijsresearch*, Lisse: Swets en Zeitlinger.
Tomic, W. (1983). *Wiskundeonderwijs: context en onderwijsactiviteiten*. Enschede: Universiteit Twente.
Tomic, W. (1985). *Docentgedrag en leerresultaten*. (Teaching behavior and student learning outcomes). Enschede: Grafoplan.
Van der Sijde, P.C. (1984). *Het leeromgevingsonderzoek: Verslag van het vooronderzoek en het ontwerp in de vertaalfase*. Enschede: Technische Hogeschool Twente, Onderafdeling der Toegepaste Onderwijskunde.
Van der Sijde, P.C. (1985a). *The implementation of a research based teacher training*. Enschede: University of Twente.
Van der Sijde, P.C. (1985b). *Classroom management als cognitief proces*. Paper presented at the ORD Enschede.
Van der Sijde, P.C. (1985c). *Het observatiesysteem TOOL-W*. Enschede: University of Twente.
Van der Sijde, P.C., (1986a). *The effect of a short teacher training course on students' learning outcomes*. Enschede: University of Twente.
Van der Sijde, P.C. (1986b). *Management en instructie in wiskundelessen (Management and instruction in math lessons)*. Enschede: Toegepaste Onderwijskunde, Universiteit van Twente.
Van der Sijde, P.C. & Tomic, W. (1987). Design of training programs for teachers based on the results of research. *Education and Society*, 37-48.
Van der Sijde, P.C. (1987). *Training the teaching script*. Almere: Versluys.
Van der Sijde, P.C. & Dirksen, A. (1987a). The TOOL-system. In Maarsse, F.J. et al.(Eds.), *Computers in psychology: methods, instumentation and psychodiagnostics*. Lisse: Swets & Zeitlinger.
Van der Sijde, P.C. & Dirksen, A. (1987b). Nieuwe ontwikkelingen in het TOOL-systeem. *Psychologie en Computers, 4*, 16-19.
Van der Sijde, P.C. (1988). Relationships of classroom climate with student learning outcomes and school climate. *Journal of Classroom Interaction, 23*, nr 2, 39-43.
Van Bemmelen, Th. & Boon, A.W. (1984). *Sigma 2MHV*. Groningen: Wolters-Noordhoff.
Van Dop, A. & Groeneveld, B. (1979). *Sigma 2HV*. Groningen: Wolters-Noordhoff.
Vonk, J. (1984). *Leraar worden, ga er maar aan staan: problemen van beginnende leraren nader bekeken*. Amsterdam: VU Boekhandel.

Walberg, H.J. (1984). Improving the productivity of American schools. *Educational Leadership, 41*, 19-27.
Walberg, H.J., Pascal, R.A. & Weinstein, T. (1985). Homework's powerful effects on learning. *Educational Leadership, 42*, 76-79.
Wardenaar, E. & Willems, J. (1984). Leren door uitleg. In J. Lowyck en H.H. Tillema, (Red.), *Vormen van leren en onderwijzen*. Lisse: Swets & Zeitlinger.